



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학석사 학위논문

칸트의 근본악 개념 연구

A Study on Kant's Concept of Radical Evil

2013년 8월

서울대학교 대학원

철학과 서양철학전공

이덕균

칸트의 근본악 개념 연구

A Study on Kant's Concept of Radical Evil

지도교수 백 종 현

이 논문을 문학석사 학위논문으로 제출함

2013년 4월

서울대학교 대학원

철학과 서양철학전공

이 덕 균

이덕균의 석사학위논문을 인준함

2013년 4월

위 원 장 이 남 인 (인)

부 위 원 장 백 종 현 (인)

위 원 김 상 환 (인)

국문초록

이 논문은 임마누엘 칸트가 『이성의 한계 안에서의 종교』에서 주장한 인간의 근본악 테제에 대해 해명하는 것을 목표로 한다. 근본악 테제란 모든 인간이 자유롭게 획득한, 현상계에서의 모든 행위의 근저에 놓여 있어 근절할 수 없는 악으로의 성벽(Hang)을 가지고 있으며, 따라서 인간은 자연 본성적으로 악하다는 것이다.

이 테제는 발표된 직후부터 칸트가 자신의 철학을 배반하고 기독교 교의와 타협했다는 의혹을 불러 일으켰으며, 따라서 진지한 철학적 논의의 대상으로 다뤄지지 않았다. 그러나 이 테제에 대한 논의가 제대로 이루어지지 않은 것은 이 테제의 정치적, 종교적 미심쩍음 때문만은 아니었다. 칸트가 인간이 근본적으로 악하다는 것에 대해서 제대로 된 증명을 제시하고 있지 않다는 것 또한 이 테제에 대한 논의를 어렵게 한 요인이었다.

그러나 최근 칸트의 후기 저작들에 대해 다양한 연구가 진행되면서 근본악 테제를 칸트의 전체 철학적 기획 속에서 중요한 의미를 갖는 것으로 이해하려는 시도가 이루어지고 있다. 이러한 시도들은 우선 근본악 테제에 대한 증명을 재구성하는 것을 목표로 하는데, 이 증명들은 크게 두 가지로 나뉜다. 하나는 ‘초월적’ 증명이며, 다른 하나는 ‘인간학적’ 증명이다. 전자는 인간의 보편적 악성에 대한 칸트의 주장을 선험적 종합 판단으로 보고, 이에 대한 초월적 논변을 제시하려 한다. 후자는 칸트가 인간의 본성에 대해 논한 다양한 텍스트를 실마리로 삼아 인간의 본성 안에 인간을 악하게 하는 요소가 있음을 보이려 한다.

그러나 이 두 종류의 시도들은 모두 인간의 근본적인 악성을 인간 자신이 어찌할 수 없는 것에서 찾고 있다는 점에서 근본악 테제에 대한 진정한 증명이라고 할 수 없다. 인간이 악하다고 하려

면 그가 그 악에 대해 책임이 있어야하기 때문이다.

본 논문은 악의 귀책성을 설명할 수 있는 근본악 테제에 대한 또 다른 증명을 제시하는 것을 목표로 한다. 이를 위해 본고는 다른 논의에서는 중요하게 다뤄지지 않았던 ‘악한 마음씨’ 개념과 ‘악으로의 성벽’ 개념의 차이에 주목하였다. 이 두 개념은 모두 도덕적 동기보다 자기사랑의 동기를 우위에 두는 인간의 악을 가리키는 개념으로서, 전자는 개인의 도덕성을 논할 때 사용되고 있으며, 후자는 인류의 도덕성을 논할 때 사용되고 있을 뿐 『종교』에는 두 개념의 차이에 대한 설명은 존재하지 않는다. 하지만 표면적으로는 유사한 이 두 개념은 그 세부적인 내용에서는 매우 큰 차이를 드러낸다. ‘악한 마음씨’에서는 인간의 악함이 더 이상 그 근거를 물을 수 없는 단 한 번의 도덕적 선택의 결과라고 설명되는 데 반해 ‘악으로의 성벽’에서는 인간의 악이 단계적으로 심화되는 것으로 서술하고 있기 때문이다. 따라서 이 두 개념이 각각 개인과 인류를 논의 대상으로 하여 동일한 도덕적 사태를 다루고 있다고 볼 수 없으며, 오히려 완전히 다른 도덕적 사태를 다루고 있다고 이해해야 한다. 논자는 칸트가 도덕적 악의 본성을 자신의 도덕성에 대한 자기기만으로 이해하고 있다는 점에 착안하여 ‘악한 마음씨’가 인간의 악에 대한 본래적 설명이며, ‘악으로의 성벽’은 ‘악한 마음씨’를 가진 인간이 자기기만의 결과 갖게 된 자신의 도덕성에 대한 잘못된 표상이라고 해석하였다.

이처럼 ‘악으로의 성벽’을 자신이 선하다는 인간의 기만적인 자기 이해에서 시작되는 것으로 봄으로써, 모든 인간이 이러한 성벽을 갖고 있다는 것, 즉, 악하다고 보아야 하는 이유 또한 해명된다. 악이 자신을 선하다고 믿게 하는 자기기만의 형태로 나타난다면, 악을 방지하기 위해서 우리는 자신을 악하다고 생각해야 하며, 이 악의 책임이 전적으로 자신에게 있다고 이해해야 한다. 이러한 방식으로 악의 귀책성 문제는 해소되며 악의 보편성은 간접적인 방식으로 증명된다. 악이 만연한 세상에서 선하기 위해 노력하는 그

누구도 악의 보편성 주장에 맞서 자신 또는 다른 누군가의 선함을 반례로 들 수 없기 때문이다.

주제어 : 근본악, 도덕성, 도덕적 가치, 악의 귀책성, 악한 마음씨, 악으로의 성벽, 자기기만

학 번 : 2009-22773

목 차

I. 서론	1
II. 본론	2
1. 『종교』에 나타난 인류의 악성	2
1.1. 인간의 자연 본성적 도덕적 성질의 의미	2
1.2. 악의 가능성	5
1.3. 마음씨(Gesinnung), 소질(Anlage), 성벽(Hang)	10
1.4. 악의 보편성에 대한 실패한 증명	21
2. 악의 보편성에 대한 두 가지 대안적 증명	25
2.1. 악의 보편성에 대한 엘리슨의 초월적 연역	26
2.2. 우드의 인간학적 증명	29
3. 악의 보편성에 대한 또 다른 증명	38
3.1. 악한 마음씨와 악으로의 성벽	40
3.2. 악의 보편성에 대한 요청	46
III. 결론	48
참고문헌	51
Abstract	54

I. 서론

칸트는 『이성의 한계 안에서의 종교』(이하 『종교』)에서 인간이 자연본성적으로 악하다는 주장을 한다. 인간은 악으로의 성벽(Hang)을 가지고 있는데, 이 성벽은 인간 자신이 자유롭게 획득한 것이며, 이 성벽은 인간이 현상계에서 하는 모든 행위의 근저에 놓여 있는 것이므로 근절될 수 없다는 것이다. 따라서 근절될 수 없으며, 인간의 모든 행위를 악하게 하는 성벽을 가지고 있는 존재인 인간은 근본적으로 악하다는 것이다. 이와 같은 주장은 칸트를 계몽 사상가로 받아들이고 있던 당대의 수많은 지식인들에게 기독교 전통과의 타협으로 비춰 졌으며, 따라서 진지한 철학적 논의 대상이라 여겨지지 않았다.¹⁾ 그 후 현재에 이르기까지 칸트의 근본악 테제는 많은 경우 철학적 동기를 결여한 것으로 취급받거나 칸트의 종교철학의 일부로 다루어졌다.²⁾ 그러나 던컨이 지적하고 있듯이 『종교』에서 칸트가 전개한 근본악 테제가 자신의 철학을 기독교 전통의 언어로 옮긴 데 불과하거나 기독교의 중요 개념을 자신의 철학을 통해 이해해 보려고 시도한 것이라면 『종교』는 그 자체로는 전혀 중요한 논의가 아니며, 우리가 칸트의 철학을 이해하기 위해 이 텍스트를 진지하게 읽어야 할 이유는 전혀 없을 것이다.³⁾ 칸트의 철학적 기획 전체와 관련해 근본악 테제를 진지하게 다루려는 노력은 비교적 최근에야 시작되었고, 그리하여 『종교』에 담긴 철학적 모티브들과 『종교』의 독해를 통해 얻을 수 있는 철학 및 칸트의 체계에 대한 탐구가 다각도에서 행해지게 되었다. 본 논문은 이러한 새로운 시도들을 검토하

1) 설러는 『종교』가 자신에게 분노를 불러 일으켰다고 쓰고 있으며, 피테는 칸트가 기독교의 비위를 맞추기 위해 이러한 저술을 하였다고 평가한 바 있다. Fackenheim(1954), p. 340. Allison(1990), p. 146 참조.

2) Wood(1970)는 칸트의 근본악 테제가 인간의 도덕적 완성을 위해 필요한 도덕적 신앙에 관한 논의를 발전시키는 과정에서 나왔다고 주장하고 있다. 신옥희(1982) 또한 칸트가 『종교』에서 근본악 테제를 통해 도덕적 신 존재 증명을 시도한 것으로 해석하고 있다. O'Neill(1989)과 Hare(1990)는 칸트의 근본악 테제가 기독교의 신학적 내용을 고쳐 쓴 것에 불과하다고 해석하고 있으며, 특히 Hare는 칸트의 논의가 기독교 신학의 몇 가지 중요 개념을 생략하였기 때문에 불완전한 것으로 평가하고 있다.

3) Duncan(2011), p. 5 참조.

여 『종교』를 통해 칸트의 실천 철학에 접근해 보려는 시도이다.

『종교』에 서술된 칸트의 악에 대한 논의는 그리 길지 않음에도 불구하고 그 안에는 많은 해석상의 어려움이 존재한다. 하나는 인간의 악성에 대한 논의에 일관성이 결여되어 있다는 것이며, 다른 하나는 인간의 악성의 보편성에 대한 증명이 제대로 전개되지 않았다는 것이다. 이 논문은 이 두 가지 문제를 해결하는 것을 목표로 한다. 먼저, 인간의 악성에 대한 칸트의 혼란스런 서술들을 재구성하여 인간의 악성을 일관성 있게 설명할 것이고, 다음으로 칸트를 대신해 제시된 악의 보편성에 대한 증명들을 소개 및 비판하고 악의 보편성 증명 문제를 다루기 위한 대안을 소개할 것이다. 이를 통해 칸트 실천철학의 핵심적 주장을 견지하면서 텍스트의 세부 사항을 생략하지 않은 근본악 테제 해석이 가능해 질 것이다.

II. 본론

1. 『종교』에 나타난 인류의 악성

1.1. 인간의 자연본성적 도덕적 성질의 의미

인간이 자연본성적으로 선한지 악한지 하는 물음을 다루면서 칸트는 우선 이 물음에 절충적인 대답은 존재하지 않음을 분명히 한다. 인간은 자연본성적으로 선하지도 악하지도 않다거나 선한 동시에 악하다는 주장은 불가능하다는 것이다. 왜냐하면 인간 본성의 선함 또는 악함이란 한 개인이 그에 따라 행동하기를 선택하는 모든 구체적인 준칙들을 결정하는 기본 준칙에 따라 결정되기 때문이다. 즉, 인간은 도덕법칙에 대한 존경을 행위의 동기로 삼는다는 원칙을 자신의 모든 특정한 준칙을 결정하는 기본 준칙으로 채택 했는가 아닌가에 따라 선하거나 악하다고 판정되는 것이니 말이다. 따라서 인간은 선한 동시에 악할 수 없다. 왜냐하면 이 기본 준칙은 인간의 행위를 결정하는 각각의 모든 준칙들을 결정하는 일

반 준칙이기 때문이다. 또한 인간은 선하지도 악하지도 않을 수 없다. 왜냐하면 인간은 준칙이 없이는 아무런 행동도 할 수 없으므로 인간은 선하지도 악하지도 않을 수 없다. 따라서 칸트가 인간의 자연본성적 도덕적 성질에 대해 말할 때 그것은 “자유로부터의 행위들의 근거와 반대되는 것”을 의미하는 것이 아니라 “눈에 띄는 모든 행실에 선행하는, 단지 (객관적 도덕법칙들 아래에서) 자유 일반을 사용하는 주관적 근거를 의미하는 것”(6:20-21) 뿐이다.

그러므로 인간의 자연본성적 도덕적 성질이란 인간의 선택에 열려 있는 것이다. 그렇지 않으면 개인은 기본 준칙을 선택한 데 대해 책임을 갖지 않을 것이며, 도덕적 선과 도덕적 악이라는 개념 일반이 의미가 없게 될 것이다. 그러므로 “악의 근거는 경향성에 의해 의사를 규정하는 객관 안에, 자연적 추동 안에 있는 것이 아니라, 의사가 자기의 자유 사용을 위해 스스로 정하는 규칙, 다시 말해 준칙에 놓여 있다.”(6:21) 그러므로 인간이 자연본성적으로 악하다고 할 때, 이는 인간에게 다르게 될 가능성이 없었다는 것을 의미하지 않는다. 실제로 인간이 다른 가능성을 취하지 않는다 하더라도 인간이 도덕법칙을 자신의 행위의 동기로 삼는 기본 준칙을 택하는 것은 원리상 가능한 것이다.

그런데 이 기본 준칙의 채택에 대해서는 다시금 어떤 근거로 이것을 채택하게 되었는지 물을 수 없다. 왜냐하면 “이 근거가 최종적으로 그 자신 더 이상 준칙이 아니고, 순전한 자연적 추동이라면, 자유의 사용은 전적으로 자연 원인들에 의한 규정들로 환원될 수 있을 터인데, 그러나 그런 일은 자유에 모순되기 때문이다.”(6:21) 따라서 “인간이 [자연]본성적으로 선하다 또는 인간이 [자연]본성적으로 악하다는 것은, 인간이 선한 준칙을 채택하는, 또는 악한(법칙에 반하는) 준칙을 채택하는 (우리에게는 탐구 불가능한) 제일의 근거를 함유하고 있고, 그것도 인간으로서 보편적으로 함유하고 있으며, 그러니까 인간은 이것을 통해 동시에 그의 유(類)의 성격을 표현하고 있다는 것을 의미한다.”(6:21) 즉, 인간은 자신의 도덕적 성격을 스스로 창시하나 인간은 그와 같은 창시의 원인에 대해 그것이 자기 자신에게 책임이 있는 것이라는 것과 이 원인

이 모든 인간 안에 존재한다는 것 외에 아는 바가 아무 것도 없다. 이와 같이 인간의 자연본성적 도덕적 성질의 최종적 의미가 밝혀진다. 그것은 인간이 실제로 경험 세계에서 행하는 모든 행위의 주관적 근거라는 점에서 자연본성적이며, 또한 유적 관점에서 본 인간 전체에게 부여될 수 있는 것이라는 점에서 자연본성적이다.

그러나 인간의 자연본성적 도덕적 성질에 대한 이런 설명에는 많은 난점이 존재한다. 칸트가 자연본성적이라는 말에 두 가지 의미를 부여하고 있기 때문이다. 하나는 우리 자신에 의해 채택된 자유 사용의 궁극적 근거라는 의미이며, 다른 하나는 유적 관점에서 본 인간에게 타당하다는 의미이다. 전자의 의미를 따를 경우 우리는 자연본성적이라는 말이 갖는 ‘실제 행위 이전에 결정되어 있음’이란 의미를 살림과 동시에 도덕을 논하는 데 있어 결코 포기될 수 없는 귀책 가능성 또한 유지할 수 있다. 그러나 후자의 의미를 따를 경우 귀책 가능성의 문제를 규명하기가 어려워진다. 인간이 유적 존재로서 필연적으로 특정한 도덕적 성질을 갖는다면, 한 개인은 그 성질이 아닌 다른 성질을 선택할 수 없을 것이고, 그렇다면 그 성질은 자유롭게 선택된 것이 아니며 따라서 그에게 책임이 물어질 수 없는 것일 것이다. 반대로 귀책 가능성을 유지하기 위해 모든 도덕적 성질을 개인의 우연적인 선택으로 환원시킨다면 모든 인간이 자유롭게 내린 도덕적 선택이 동일한 결과를 낳을 가능성에 대한 증명이 필요할 것이다. 그러므로 칸트가 독자적인 의미를 부여해 사용하고 있는 ‘자연본성적 도덕적 성질’은 그 시작부터 해명해야 할 난점을 가지고 있다고 할 수 있다. 첫 번째 의미로 사용될 경우 그것은 아무 어려움도 불러일으키지 않는다. 그러나 두 번째 의미와 결합될 때 그것은 칸트의 ‘근본악’ тезис의 설득력 자체를 위협하는 근본적 난점을 발생시키게 된다. 이 난점이 무엇인지 분명히 이해하기 위해서는 칸트가 『종교』에서 사용하고 있는 여러 개념들에 대한 이해가 필요하므로, 아래에서 칸트의 개념을 소개하며 난점을 구체화시켜 보도록 하자.

1.2. 악의 가능성

칸트의 철학에서 인간의 악은 쉽게 이해하기 어려운 문제이다. 칸트에게서 도덕성은 자유 및 이성과 밀접히 연결되어 있기 때문이다. 칸트의 실천철학의 핵심을 이루는 생각은 자유로운 행위는 도덕법칙에 대한 존경심에 의해 동기화된 행위라는 것이다. 이는 『실천이성비판』의 다음 구절에서 가장 잘 나타나 있다.

“저 보편적인 법칙 수립적 형식 이외에는 이 사건들에 대한 어떠한 규정 근거도 법칙으로 쓰일 수 없다면, 그러한 의지는 현상들의 자연 법칙, 곧 현상들 상호간의 인과 법칙과는 전적으로 독립적인 것으로 생각되어야 한다. 그러한 독립성은 그러나 가장 엄밀한, 다시 말해 초월적 의미에서 자유라 일컫는다. 그러므로, 준칙의 순전한 법칙 수립적 형식이 그에 대해 오로지 법칙으로 쓰일 수 있는 의지는 자유 의지이다.”(5:29)

또한 『윤리형이상학 정초』(이하 『정초』)에서 칸트는 자유 의지에 대해 다음과 같이 쓰고 있다.

“의지의 자유가 자율, 다시 말해 자기 자신에게 법칙인 의지의 성질 말고 다른 무엇일 수 있겠는가? 그러나 ‘의지는 모든 행위에 있어 자기 자신에게 법칙이다’라는 명제는 단지, 자기 자신을 또한 보편적 법칙으로서 대상으로 가질 수 있는 준칙 외의 다른 어떤 준칙에 따라서는 행위하지 않는다는 원리를 표시할 따름이다. 그러나 이것은 바로 정언 명령의 정식이자 윤리성의 원리이다. 그러므로 자유 의지와 윤리 법칙 아래에 있는 의지는 한 가지이다.”(6:447)

여기서 자유 의지는 법칙을 따르는 의지와 동일시되고 있다. 그러나 이는 시지윅(H. Sidgwick)의 다음과 같은 유명한 비판에서 잘 나타나는 문제를 발생시킨다. 시지윅에 따르면 칸트의 철학에서 어떤 행위는 그것이 도덕적으로 용인 가능한 것일 경우에만 이성적이고, 그렇기에 자유롭다. 그렇다면 행위의 비도덕성이란 그 행위가 자연 인과성에 의해 결정

된 것이란 의미일 것이며, 그 때문에 행위자에게 책임을 물을 수 있는 것이 아니다. 따라서 칸트의 철학에서 책임을 물을 수 있는 비도덕적인 행위는 논리적으로 불가능하다.⁴⁾ 칸트가 『정초』와 『실천이성비판』에서 악의 의미에 대해 적극적인 설명을 하지 않고 있는 것은 도덕적-악의 불가능성에 대한 이러한 의심을 더욱 강화시킨다. 칸트는 『실천이성비판』에서 도덕적 악과 자연적 악을 구분하면서 실천철학의 논의 대상이 될 수 있는 악은 인간 자신에 의해 초래된 도덕적 악 뿐임을 분명히 하고 있으나⁵⁾ 도덕적 악 그 자체의 성질에 대해서는 아무런 설명도 하고 있지 않다. 또한 『정초』에서도 의무의 원리에서 벗어난 행위를 악으로 규정하고,⁶⁾ 의무에서 벗어나는 많은 행위의 실례들을 들고 있으나 인간이 자유롭게 악을 범할 수 있는 가능성에 대한 논의는 하고 있지 않다.

이렇듯 칸트의 비판기 초기 실천철학에는 악의 문제를 분명히 다루기 위한 개념 장치가 결여되어 있다. 이는 이 두 저서에서 칸트가 선이란 무엇인가란 문제를 해명하는데 초점을 맞추고 있기 때문일 것이다. 그러나 악을 본격적으로 주제화하는 『종교』에서는 이를 다루기 위한 개념이 도입되어야만 한다. 『종교』에서 새로이 도입되는 개념 장치는 다음 두 가지이다. 먼저, 『종교』는 앞의 두 실천철학 저서들과는 다르게 ‘선행’이 무엇인가가 아니라⁷⁾ ‘선한 사람’은 어떤 사람인가라는 물음을 던진다⁸⁾. 즉, 행위의 도덕성이 아니라 행위자의 도덕성이 주제로 다뤄

4) Sidgwick(1888), pp. 405-412.

5) “우리가 한 행위에서 그 행위의 **선**, 악을 고찰하느냐, 우리의 복, 화(해악)를 고찰하느냐는 두 가지 전혀 별개의 관정의 문제다.”(5:59-60)

6) “내가 만약 의무의 원리에서 벗어난다면, 그것은 전적으로 확실히 악하다.”(4:403)

7) 『정초』를 시작하는 유명한 구절이 이를 잘 보여준다. “이 세계에서 또는 도대체가 이 세계 밖에서까지라도 아무런 제한 없이 선하다고 생각될 수 있을 것은 오로지 **선의지**뿐이다.”(5:393) 이 글에서 칸트의 주제는 선이란 무엇인가에 대한 탐구이다. 칸트는 뒤이어 수많은 개별적인 행위들을 검토하면서 어떤 행위가 선하고 어떤 행위가 악한지를 따진다. 그러나 이런 방식으로는 악한 행위와 자유 의지의 관계를 물을 수 없다. 의지는 행위가 아니라 행위자에게 속하는 것이기 때문이다. 따라서 행위자가 악한 행위를 하는 것이 어떻게 가능한가를 묻지 않으면 자유와 악 사이의 관계는 밝힐 수 없는 것이다.

8) 『종교』를 여는 질문은 인류의 역사가 (도덕적)발전의 역사였는지 퇴보의 역사였는지 이다. 두 견해는 각각 인류의 도덕적 선함과 도덕적 악함을 전제한다. 이로부터 칸트는

지는 것이다. 이로써 자유의지를 갖는 행위자가 어떻게 도덕법칙에 어긋나는 행위를 하는가라는 문제를 해명할 여지가 생긴다. 다음으로, 『종교』에서 칸트는 ‘의지(Wille)’와 ‘의사(Willkür)’를 분명히 구별해 사용한다. 『종교』에서 칸트는 ‘의지’라는 표현을 단 한 차례만, 그것도 인간의 의지가 아니라 ‘신적의지’에 대해 논할 때 사용하고 있으며 인간의 행위 및 그 자유를 논하는 데는 언제나 ‘의사’라는 표현을 사용한다. 이로써 법칙을 따르지 않는 인간의 자유에 대해서 논할 수 있는 여지가 생긴다. 이 두 개념 장치가 문제 해결에 각각 어떻게 기여하는지를 아래에서 살펴보자.

칸트의 『종교』는 개별적인 행동의 도덕성이 아니라 행위자의 도덕성에 대한 설명을 제공함으로써 악의 가능성 문제를 해명하고 있다. 시지윅의 비판은 행위의 도덕성에 초점을 맞추어 이루어진다. 선한 행위가 자유로운 행위라면 악한 행위는 자유롭지 않은 행위이며, 따라서 우리에게 책임이 없는 행위이다. 『정초』의 논의들은 행위의 도덕적 성격에 초점이 맞추어져 있으므로 우리가 행위자에 대해 고려하지 않고 칸트의 실천철학을 읽으면 시지윅이 제기한 것과 같은 의문이 자연스럽게 떠오른다. 그러나 행위가 아니라 행위자에 초점을 맞추면 상황은 달라진다. 우리는 행위자에 초점을 맞추어 비도덕적인 행위를 다음과 같이 재구성해 볼 수 있다.

- 1) 우리는 도덕법칙에 대한 존경이라는 동기와 자연적인 경향성에서 오는 다른 동기들을 행위를 위한 동기들로 가지고 있다.
- 2) 우리가 도덕법칙에 대한 존경이 아닌 자연적인 경향성을 동기로 삼을 것을 자유롭게 결정한다.
- 3) 그 때 이 행위는 자연적인 사건의 계열에 속하는 자유롭지 않은 행위이다.

그렇다면 도덕적 악은 도덕법칙이 아닌 다른 것들을 행위의 동기로 채

인간의 도덕적 성질에 대한 논의로 나아간다.

택하는 때 선택들에 있는 것이지, 도덕법칙에 대한 존경을 동기로 하지 않는 각각의 행위들에 있는 것이 아니다. 즉, 도덕법칙에 따라 일어나지 않은 각각의 행위들은 자유롭지 않은 행위이며, 자연적인 행위로서 행위 그 자체는 도덕적 평가의 대상이 아니다. 도덕적 악이라고 평가 받을 수 있는 것은 오직 도덕법칙을 동기로 선택하지 않은 행위자의 선택, 자유롭지 않은 행위를 할 것에 대한 자유로운 선택뿐이다. 따라서 도덕법칙에 따르지 않는 자유롭지 않은, 자연 인과성에 따라 일어난 행위라 할지라도 그 행위자에게 도덕적인 책임을 물을 수 있다. 우리는 자유로운 행위를 하지 않은 데 대해 책임이 있다. 이처럼 행위를 자유로운 것으로 만드는 것이 무엇이나는 질문을 행위자를 자유로운 것으로 만드는 것이 무엇이나는 질문에 종속시킴으로써 우리는 시지윅의 비판을 피해 칸트의 철학 내에 악을 정초할 수 있는 것이다.⁹⁾

그러나 여기서 다시금 다음과 같은 문제가 발생한다. 우리는 어떻게 자유롭지 않은 행위를 할 것을 스스로 자유롭게 선택할 수 있는가? 자유로운 의지가 법칙을 따르는 의지라면, 자유로운 의지가 어떻게 도덕법칙이 아닌 다른 것을 행위의 동기로 채택할 수 있겠는가? 이 물음에 답하기 위해 새롭게 도입되는 개념이 ‘의지’와 구분되는 ‘의사’의 자유이다.¹⁰⁾

칸트는 『종교』에서부터 이미 ‘의지’가 아니라 ‘의사’를 인간의 행위를 설명하기 위한 주요 개념으로 사용하고 있지만, 이 두 개념의 분명한 구분이 이루어진 것은 보다 뒤에 나온 저서인 『윤리형이상학』(1798)에 서이다. 이에 따르면 ‘의지’와 ‘의사’는 모두 “욕구능력(Begehrungsvermögen)이나 ‘의사’는 “대상을 산출할 수 있는 자신의 행위 능력에 대한 의식”과 관련된 것이고, ‘의지’는 “욕구 능력의 내적 규정근거”, “주관의 이성”과 관련된 것이다. “따라서 의지는 (의사처럼) 행위와의 관계에서가 아니라 오히려 행위와 관련된 의사를 규정하는 근거와 관련해서 고려되는 행위 능력이다. 그리고 이 의지는 본래 자신에 앞서 어떠한 규정 근거도 가지고 있지 않으며, 오히려 이 의지가 의사를 규정할 수 있는 한에서, 실천이성 자체이다.”(MS, RL, 213) 의지는 의

9) Sussman, 155

10) 이하 ‘의지’와 ‘의사’의 구분에 대해서는 김종식(2004)의 논문을 참고로 하였다.

사를 규정한다. 그리고 의지는 실천이성과 동일한 것이다. 그러므로 의지는 객관적이고, 보편성을 갖는 것으로서 법칙을 산출한다. 그런데 의사는 또한 경향성의 영향도 받는다. 의지와 경향성 양 측의 영향을 모두 받는 의사는 주관적인 것이며, 의사는 준칙을 산출한다.¹¹⁾ 따라서 의지와 의사 중 자유롭다고 할 수 있는 것은 의사뿐이다.¹²⁾ 왜냐하면 법칙에 관계하는 의지는 그 자체로 필연적인 것으로서 자유롭다거나 자유롭지 않다거나 할 수 없기 때문이다.¹³⁾ 그러나 의사는 준칙과 관계하기 때문에 이성에 의해서 영향을 받기도 하지만, 경향성에 의해 영향을 받음으로써 언제나 자유로운 선택의 기로에 놓여 있다. 따라서 의사만이 자유롭다.

이 구도에서 자유는 더 이상 법칙에 따르는 필연성과 결합되어 있지 않다. 의지는 자유라는 술어와 결합되기에는 부적절한 개념으로 다시 주조되며, 의사는 법칙에 따를 것인가, 따르지 않을 것인가라는 열린 선택에 직면해 있다는 의미에서 자유롭다.¹⁴⁾ 그리고 법칙에 따르지 않을 것을 선택한 의사는 자의로 그것을 선택한 것이므로 도덕적으로 악하다고 할 수 있다. 그리고 의사는 의지와는 다르게 도덕법칙뿐만 아니라 경향성도 자신의 동기로 채택할 수 있다. 바로 그렇기 때문에 의사는 자의로 도덕법칙에서 이탈할 수 있는 것이며, 그런 한에서 인간은 도덕적으로 악할 수 있는 것이다.

그런데 주의해야 할 것은 인간의 도덕적 악이 행위가 일어나는 순간마다 내려지는 의사의 선택에 의해 평가되는 것이 아니라는 것이다. 만약 그렇다면 인간은 어떤 순간에는 선하며, 어떤 순간에는 악해서 총체적으로 인간은 선한 동시에 악한 존재라고 평가될 수 있을 것이다. 그러나 이는 칸트의 도덕적 엄격주의에 위배된다. 게다가 인간이 매 순간 서로

11) MS, RL, 226

12) Ibid.

13) MS, RL, 226

14) 김종식은 의지=실천이성에 따르는 의사만이 진정한 의미에서 자유로우며, 선택에 열려 있는 의사의 자유는 넓은 의미의 자유로서, 칸트가 자유를 포괄적인 의미로 사용하고 있음을 지적하고 있다. 자유 개념의 이러한 포괄적 사용이 과연 타당한지에 대해 논란의 여지가 있을 수 있으나 이 글의 목적은 칸트가 도덕적 악에 대한 논의를 소개하는 것일 뿐이므로 이에 대한 논의는 생략하기로 한다.

다른 선택을 내린다는 생각은 인간의 모든 행위를 어떤 법칙성도 찾을 수 없는 임의적인 것으로 만들 것이며, 그렇다면 우리는 인간의 도덕적 성질에 대해서 어떠한 판단도 내릴 수 없게 될 것이다. 그러나 칸트는 인간의 자유가 아무런 근거도 없이 임의적으로 행사되는 것이라고 생각하지 않는다. 『정초』에서 칸트는 자유를 “법칙들에 따르는 원인성”(4:446)이라고 정의하고 있거니와 『종교』에서도 “자유 일반을 사용하는 주관적 근거”(6:21)에 대한 탐구를 연구 대상에 포함시키고 있다. 그러므로 인간의 도덕적 성격에 대해 판단을 내리기 위해서는 인간의 모든 행위와 이 행위를 이끄는 각각의 준칙들을 최종적으로 자유롭게 결정하는 궁극적인 근거를 묻지 않을 수 없으며, 그것은 다시금 자유로운 것이므로 이 근거는 의사에 의해 자유롭게 선택된 것이어야만 한다. 이 의사가 우리의 모든 준칙과 이에 따르는 행위들을 결정하므로 이 의사의 성격이 우리의 도덕적 성격을 결정한다. 의사가 도덕법칙을 따르도록 우리를 규정하면, 우리는 선하고 그렇지 않다면 악하다. 그리고 의사가 우리의 자유 사용의 궁극적 근거를 악한 것으로 택하였다면 우리의 모든 행위 또한 악할 수밖에 없다. 칸트는 『종교』에서 의사에 의해 선택된 자유 사용의 제일의 근거를 마음씨(Gesinnung)라 부르고 있다. 이렇게 시지윅이 제기했던 귀책 가능한 악의 가능성 문제가 해결된다. 자유롭게 지 않은 행위를 하게 만드는 바로 그 마음씨에 귀책 가능한 악이 놓여 있다.

1.3. 마음씨(Gesinnung), 소질(Anlage), 성벽(Hang)

1.3.1. 악한 마음씨와 선한 마음씨

마음씨는 인간의 선하거나 악한 자유사용의 제일근거로서 인간의 도덕적 성격을 결정하는 것이다. 마음씨가 인간의 도덕적 성격을 결정한다는 것이 어떠한 사태인지를 보다 분명히 이해하기 위해 이제 인간의 마음씨

와 인간의 여러 행위들이 어떤 관계를 맺고 있는지를 살펴보도록 하자.

『종교』에서 칸트는 우리의 행위가 어떻게 이루어지는지에 대해 상세한 설명을 제시하고 있는데, 이에 따르면 악한 마음씨에서 선한 행위는 나올 수 없다. 그것은 다음과 같은 이유 때문이다.

(1) 인간은 준칙에 따라 행위한다.

(2) 그런데 인간의 준칙은 어떤 목적과 그것을 이루기 위한 수단으로 이루어진다. 그리고 인간이 자신의 준칙의 목적으로 취할 수 있는 것, ‘동기(Triebsfeder)’는 두 가지 뿐이다. 하나는 도덕법칙으로, “도덕법칙은 그 자체만으로, 이성의 판단에서, 동기가 되며, [...] 법칙이 그 법칙과 관련되는 행위와 관련하여 누군가의 의사를 규정하고 있지 않다면, 법칙과 대립하는 어떤 동기가 그의 의사에 영향을 미치고 있는 것임이 틀림 없다”(6:23) 이 다른 하나의 동기는 인간이 **자연소질(Anlage)**로 말미암아 갖고 있는 감성의 동기이며, 자기사랑의 원리에 따라 동기로서 준칙에 채용되는 것이다(6:36).

(3) 그런데 자유사용의 제일의 근거는 그 자신이 또한 하나의 준칙으로서, 바로 그에 의해 다른 모든 준칙들이 결정되는 최상의 준칙이다(6:31). 그렇다면 우리의 도덕적 성격이란 우리가 최상의 준칙으로서 도덕법칙을 채용했는가, 아니면 자기사랑의 동기를 채용했는가에 따라 결정되는 것이다.

(4) 그리고 이 최상의 준칙에서 다른 모든 준칙들이 따라 나오므로 한번 도덕법칙 준수를 자신의 최상의 준칙으로 선택한 사람이 다시금 자기사랑의 준칙을 택하는 일이나 또는 그 반대의 일은 그가 동일한 사람으로 남아 있는 한 불가능하다. 그와 같은 일은 칸트가 ‘마음씨의 혁명’이라고 부를 수 있는 사건에 해당할 터인데, 이와 같은 혁명을 거친 인간을 그 전과 동일한 도덕적 성격을 가진 인간이라고 부를 수는 없을 것이

다.

그러므로 일단 우리가 어떤 인간을 도덕적으로 선하거나 악하다고 판단하면 우리는 그가 경험 중에 행하는 모든 행위를 악한 것으로-합법칙적이든 반법칙적이든- 상정해야 한다. 우리가 악한 마음씨를 가지고 있다면 이 세계에서 행한 우리의 모든 행위는 악할 것이며, 반대라면 우리의 모든 행위는 선할 것이다. 이로써 칸트가 자연본성적 도덕적 성질이라는 말로써 의미했던 바가 좀 더 분명히 밝혀진다. 우리는 현상계에서의 우리의 모든 행위에 선행하여 그것들을 결정하는 최상의 준칙을 자유로운 선택의 결과로 갖고 있다. 바로 이러한 의미에서 인간은 자연본성적 도덕적 성질을 가지고 있다. 그것은 “출생이 바로 선, 악의 원인이라는 그러한 의미가 아니다.”(6:32)¹⁵⁾

이제 우리는 인간의 행동을 그것을 결정하는 마음씨와 관련하여 다음과 같이 정리해 볼 수 있다. 인간은 반법칙적으로 행동하거나 합법칙적으로 행동한다. 그리고 합법칙적 행위는 도덕법칙에 대한 존경심 한 행위와 자기 이익을 위해 한 행위로 구별된다. 반법칙적인 행위의 경우 그 행위자는 악한 마음씨를 가지고 있는 것이다. 합법칙적인 행위의 경우 그 행위자는 악한 마음씨를 가지고 있을 수도 있고, 선한 마음씨를 가지고 있을 수도 있다. 합법칙적인 행위를 한 사람의 마음씨를 판단하기 위해 우리는 그가 태어나서 한 일련의 모든 행위를 근거로 사용할 수 있다. 선한 마음씨에서는 단 하나의 악한 행위도 나올 수 없기 때문에 그가 단 한 번이라도 반법칙적 행위를 했다면 그는 악한 마음씨를 가지고 있는 것이며 그의 수많은 합법칙적 행위들은 오직 이기적인 동기에서만 행해진 것이다. 그런데 태어나서 단 한 번도 법칙을 어기지 않고 오직 합법

15) 자연본성적 도덕적 성질에 대한 이와 같은 이해에 따라 우리는 “단 하나의 의식적인 악한 행위로부터 그 행위의 기초에 놓여 있는 악한 준칙을 추론해 내야만 한다.”(6:20) 이것은 우리가 악이라는 표현을 사용하는 일반적인 방식에 어긋난다. 우리가 어떤 사람을 악하다고 할 때는 그가 상당한 정도의 비행을 저질렀음을 의미하기 때문이다. 때문에 악에 대한 칸트의 논의는 악에 대한 우리의 일상적 이해와 크게 상충한다는 단점을 지니고 있다. 그러나 최소한 칸트의 설명은 칸트의 체계 내부에서는 타당하다. 칸트는 도덕법칙에 대한 존경에서 나온 행위만을 유일하게 선한 행위로 보고 있기 때문이다.

칙적인 행위만을 해 온 사람이라고 하더라도 우리는 그가 선한 마음씨를 가지고 있다고 판단할 수 없다. 아래에서 보다 상세히 논할 것이지만, 우리는 현상계 너머에 있는 행위자의 최상의 준칙을 직접 인식할 수는 없기 때문이다.

그렇다면 인간의 도덕적 성격에 대한 탐구는 결국 현상계를 예지계로부터 분리시키는 인식의 벽에 가로막혀 판단 불가능한 상태로 남겨지고 마는가? 그렇지 않다. 칸트는 논의의 초점을 예지적인 것에서 현상적인 것으로 이동시켜 탐구를 계속해 나간다.

1.3.2. 인간 자연본성 안의 선의 근원적 소질(Anlage)

윗 절에서 최상의 준칙의 동기에 대해 논하면서 칸트는 인간 안에 보편적으로 놓여 있는 ‘소질(Anlage)’이라고 하는 일종의 인류학적인 개념을 도입하고 있다. 인간의 마음씨는 인간이 도덕법칙과 맺고 있는 관계를 중심으로 논해질 수 있지만 또한 인간이 보편적으로 가지고 있는 이 자연적 ‘소질’과의 관계를 중심으로 논의될 수 있다.

‘소질’이란 “그 존재자에게 필요한 구성요소들을 뜻할 뿐만 아니라, 또한 그러한 존재자이기 위한 그 구성요소들의 결합의 형식들을 뜻한다.”(6:28) 그리고 “소질들이 어떤 그러한 존재자의 가능성에 필연적으로 속한다면 그것들은 근원적인 것이고 [...] 그 존재자가 그러한 소질 없이도 자체로 가능할 것이라면, 그 소질들은 우연적인 것이다.”(6:28) 칸트에 따르면 인간은 “선의 근원적 소질Anlage des Guten”을 가지고 있는데, 소극적으로(도덕법칙과 상충하지 않음) 선할 뿐만 아니라 적극적으로(도덕법칙의 준수를 촉진한다) 선하기도 한¹⁶⁾ 이 소질들은 다음

16) 그러나 과연 이 각각의 소질들이 어떻게 적극적으로 선을 촉진하는데 기여하는가에 대해 칸트는 분명한 설명을 제시하고 있지 않다. 오히려 동물성의 소질 및 인간성의 소질은 자기 사랑으로 인해 도덕법칙을 따르는 것을 저해한다는 의미에서 악의 근원이기조차 한 것으로 보인다. 그러나 칸트는 이 소질들은 선의 소극적, 적극적 소질로 칭함으로써 악의 근원을 이 두 자기 사랑의 소질에서 찾는 것을 명백히 거부하고 있다.

세 가지이다.¹⁷⁾

(1) 인간은 생명체로서 동물성의 소질을 가지고 있다. 이것은 물리적이고 기계적인 자기 사랑, 즉, 이성이 필요하지 않은 자기사랑이다. 이 소질은 첫째, 자기 자신의 보존의 소질이며 둘째, 종의 번식 및 성적 교접을 통해 낳은 것의 보존의 소질이고, 셋째, 다른 인간들과의 공동생활의 소질, 즉, 사회로의 충동이다. 비록 이 소질이 선의 소질이기는 하나 이로부터 패악들도 발생할 수 있다. 이것들은 자연 야성의 패악이라 할 수 있는 것으로서 최고도에 이르렀을 때 폭음폭식, 환락, 야만적 무법성이 된다.

(2) 생명체이자 이성을 가진 존재인 인간은 다음으로 인간성의 소질을 가지고 있다. 이것은 물리적이기는 하지만, 이성에 의한 비교하는 자기 사랑의 소질이다. 이에 따라 인간은 타인과의 비교에서 자신의 행복과 불행을 판정하고, 타인의 의견 중에서 하나의 가치를 얻으려는 경향성을 갖게 된다. 그 결과 인간은 어느 누구에게도 자신의 위에 서는 우월성을 허락하지 않으려고 하고, 자기가 타인 위에 서는 우월성을 획득하려는 욕구를 갖게 된다. 이 욕구가 부당함에도 불구하고 그것이 선의 소질일 수 있는 이유는 이러한 경쟁을 통해 인간이 문화를 발전시키게 되기 때문이다. 그러나 이 소질 또한 잘못될 경우 문화의 패악들을 불러 일으켜 최고로 악의적일 때 시기, 배은망덕, 남의 불행을 기뻐함 등의 패악을 낳게 된다.

(3) 마지막으로 이성적인 존재인 인간에게는 인격성의 소질이 있다. 이것은 도덕법칙에 대한 존경의 감수성, 의사의 그 자체만으로써 충분한 동기의 감수성이다. 이것은 인간이 도덕법칙에 대한 존경을 동기로써 우리 준칙 안에 채용할 주관적 근거를 가지고 있음을 의미한다. 인격성의 소질 위에는 어떠한 악도 접목될 수 없다. 동물성의 소질이 이성을 뿌리

17) 이하는 6:26-29 를 정리한 것임.

로 갖지 않고 인간성의 소질은 다른 동기들에 봉사하는 이성을 뿌리로 갖는 반면, 인격성의 소질만이 그 자체로 실천적인, 무조건적으로 법칙 수립적인 이성을 뿌리로 갖는다.

이 소질들은 인간에게 근원적인 것이므로 모든 인간은 도덕법칙에 대한 존경과 더불어 동물적이고 인간적인 자기사랑을 가지고 있다. 그리고 이것들이 인간이 자신의 도덕적 성질, 즉, 마음씨를 결정할 때 자신의 준칙 안에 채용하게 되는 두 개의 ‘동기’의 구체적 내용이다. 인간은 이들 중 어느 것도 완전히 버릴 수는 없으므로 인간의 준칙은 이 중 어느 하나만을 자신의 동기로 갖지 않는다. 인간은 이 두 동기를 모두 자신의 준칙 안에 채택하되 다만 어느 하나를 다른 하나의 조건으로 삼는 방식으로 채택한다. 도덕법칙에 대한 존경이 자기 사랑의 조건이 될 경우 인간은 선하며 그 반대의 경우 인간은 악하다. 즉, 도덕법칙에 위배되지 않는 한 자기사랑을 추구하는 것을 근본 준칙으로 삼은 경우 그 인간은 선하며 자기사랑에 위배되지 않을 경우에만 도덕법칙에 따라 행하는 것을 준칙으로 삼은 경우 그 인간은 악하다.

이처럼 선의 근원적 소질은 그 자체로는 도덕적 성격을 가지고 있지 않다. 그것은 인간이 인간이기 위한 필수적인 구성요소이므로 인간에게 책임을 물을 수 없는 것이기 때문이다. 그러나 이 소질들은 설령 동물성의 소질이라 할지라도 선의 실현에 아무런 장애도 되지 않는 것이다. 따라서 인간이 현상계의 존재자로서 자기사랑의 동기를 가지고 있고, 그에 영향을 받지 않을 수 없다는 것은 인간의 도덕적 책임을 조금도 덜어주지 않는다. 인간의 도덕적 악은 그가 자신 안에 존경의 대상으로 자리하고 있는 도덕법칙을 준수하는데 실패했기 때문에 초래된 것이 아니라 그가 자신 안에 있는 선의 소질들을 잘못된 방식으로 사용했기 때문에 초래된 것이다. 이처럼 인간의 소질들이 그 자체로는 어떠한 도덕적 어려움도 초래시키지 않는다는 것이 분명해짐으로써 자신의 도덕성에 대한 인간의 책임이 한층 분명해진다. 이로써 인간의 자연본성적 도덕적 성격의 첫 번째 의미, 즉, 자유 사용의 제일 근거로서의 자연본성은 완전히

해명된다.

그러나 여전히 인간의 도덕적 성격을 어떻게 판단해야 하는가란 문제는 해결되지 않는다. 관찰가능한 행위만으로 판단을 내릴 수 없을 때 우리는 어떻게 인간의 도덕적 성격을 판단할 것인가? 그리고 관찰은 우리에게 인간의 도덕성에 대한 불완전한 판단만을 가능하게 하는 것이라면, 인간 전체의 도덕적 성격은 또 어떻게 판단할 수 있는가? 선으로의 소질은 인간이 자신의 자유 사용을 통해 획득한 것이 아니므로 이를 근거로 인간의 선과 악을 논하는 것은 불가능하다. 따라서 인간 전체의 도덕적 성격을 논하기 위해서는 인간이 보편적으로 가지고 있는, 그러나 인간의 본성 자체에 속하는 것이 아니라 인간의 선택의 결과 획득된 무언가가 있어야 한다. 그리고 그것의 존재를 증명할 수 있다면, 개개인의 도덕적 성격으로부터 귀납하지 않고도 인류의 도덕적 성격에 대해 판단을 내릴 수 있을 것이다. 칸트는 이를 위해서 ‘성벽(Hang)’이라는 개념을 도입한다.

1.3.3. 인간 자연본성 안의 악으로의 성벽

성벽이란¹⁸⁾ 그 가능성이 인간성 일반에 대해 우연적인 경향성을 가능하게 하는 주관적 근거이다. 성벽은 선천적이라 할지라도 인간 자신에 의해 획득 또는 초래된 것으로 생각될 수 있다는 점에서 소질과는 구별된다.¹⁹⁾ 따라서 악으로의 성벽이란 인간에 의해 초래된 것이며, 도덕적

18) 이하 성벽에 대한 개괄은 6:29-32를 정리한 것이다.

19) Duncan(2011)은 칸트의 성벽을 인간의 생물학적인 특징으로 이해하고 있다. 예를 들어 인간은 특정한 물질에 반복적으로 노출되면 그에 중독=그 물질을 향한 경향성을 갖게 되는 생물학적인 성질을 가지고 있다. 우리가 알콜에 중독되는 것은 우리 안에 알콜에 중독되는 생물학적인 성질이 있기 때문이다. 그러나 그 성질을 발현시키는 것은 개인의 행위의 결과이다. 그러므로 알콜 중독은 어떤 의미에서는 우리에게 선천적인 것이되 우리에게 귀책 가능한 것이다. Duncan(2011) pp. 81-83 참조. 이 주장은 성벽과 소질의 차이를 우리의 경험에 비추어 이해 가능한 방식으로 설명하고 있다는 장점을 가지고 있다. 그러나 이는 잠재적으로 우리의 악행을 무지의 소이로 돌릴 위험을 가지고 있다. 예컨대 우리가 모르는 물질에 실수로 노출되어 중독되고, 그 결과 좋지 않은 물질에 탐닉하게 되었을 때 우리는 그에 대한 도덕적 책임을 지지 않게 된

악이 자유로운 의사의 규정으로서만 가능한 것이므로 악으로의 성벽은 준칙들을 도덕법칙에서 이탈 가능하게 하는 주관적 근거 안에 있는 것이다. 그리고 이 성벽이 인간에게 보편적으로 속한다면 이것은 인간의 악으로의 자연본성적인 성벽이다. 또한 이 성벽에서 도덕법칙을 자기의 준칙 안에 채용하거나 채용하지 않는 의사의 유능한(선한) 심정(Herz)과 무능한(악한) 심정이 생긴다.

‘성벽’에 대한 칸트의 논의의 핵심은 ‘성벽’이라는 표현이 인간에게 귀책 불가능한 어떤 자연본성적인 특징을 가리키는 것처럼 보임에도 불구하고 그것이 엄격하게 인간에게 귀책 가능한 것으로 상정되고 있다는 것이다. “어떤 자유 사용으로의 (감성적 충동에 기초하고 있는) 물리적 성벽이란 모순”이다(6:31). 따라서 악으로의 성벽은 우리 자신의 행실의 결과여야 한다. 하지만 “사람들은 성벽이란 개념을 모든 행실에 선행하는, 그러니까 그 자신 아직 행실이 아닌, 의사의 주관적 규정근거라고 이해한다.”(6:31) 칸트는 이 모순을 다음과 같은 방법으로 해결한다. 악으로의 성벽은 “최상의 준칙을 (법칙에 맞든 어긋나든) 의사 안에 채용하는 자유의 사용”이란 의미에서 인간에게 책임을 물을 수 있는 죄과인 동시에 “모든 반법칙적인 행실의 형식적 근거”라는 의미에서 모든 행실에 선행하는 것이라 할 수 있다. 따라서 이 성벽은 근절될 수가 없다. 왜냐하면 악으로의 성벽은 최상의 준칙을 법칙에 어긋나게 의사 안에 채용한 것을 의미하는데 악으로의 성벽을 근절하기 위해서는 다름 아니라 최상의 준칙이 선의 준칙이어야만 하기 때문이다. 이와 같은 방식으로 칸트는 ‘성벽’이라는 개념과 귀책 가능성을 조화시킴으로써 도덕적 악의 문제를 주제화할 수 있는 가능성을 얻게 된다. 이는 ‘성벽’에 대해 칸트가 『정초』에서 논한 바와 『종교』에서 논한 바를 비교하면 분명하게 나타난다.²⁰⁾ 『정초』에서 칸트는 성벽에 대해 다음과 같이 논하고 있다.

다. Duncan은 이를 충분한 사려를 기울이지 않은 잘못으로 규정함으로써 귀책성의 문제를 해결하려 하고 있으나(ibid. pp. 117-122) 이는 칸트의 윤리적 동기주의에 명백히 위배되는 것이다.

20) 이 논점은 Muchnik(2009) pp. 122-123을 참조로 하였다.

(경향성과 이성의 갈등에서) 의무의 저 엄격한 법칙들에 반대하여 궤변을 늘어놓고, 그 법칙들의 타당성을, 적어도 그것들의 순수성과 엄격성을 의심하여, 가능한 한 그것들을 우리의 소망이나 경향성들에 더 맞도록 만들려는, 다시 말해 법칙들을 근본적으로 변질시키고 그것들의 전 위엄을 파괴하려는 성벽이 생긴다.(4:405)

경향성과 이성 유한한 존재인 인간 안에서 행위의 원리로서 각각 동등한 타당성을 주장한다. 그 때문에 의무의 법칙들은 인간 안에서 그 효력을 잃게 된다. 경향성과 이성이 행위의 원리로서 인간 안에서 서로 대립한다는 생각은 『종교』에서 의무와 자기사랑이라는 두 가지 동기의 대립이라는 생각으로 이어지고 있다. 그러나 『정초』와 『종교』의 성벽에 대한 논의에는 다음 두 가지 차이점이 있다. 『정초』에서 악은 인간의 선택과는 무관하게 인간이 경향성을 가진 존재라는 사실에서 발생하는 것처럼 그려지고 있다.²¹⁾ 반면 『종교』에서 아담의 타락을 설명하며 칸트는 다음과 같이 성벽에 대해 논한다.

도덕법칙은 순수하지 못하고 경향성들에 유혹당하는 존재자인 인간에게서 그럴 수밖에 없듯이 금지로서 먼저 있었다. 그런데 인간은 이 법칙을 충분한 동기로 곧장 따르는 대신에, 조건적으로만 선택할 수 있는 또 다른 동기들을 둘러 찾아보았다 [...] 인간은 다른 어떠한 동기의 영향도 배제하는 계명의 엄격성을 의심하기 시작하였고, 그 다음에는 계명에 대한 순종을 수단의 한 날 (자기사랑의 원리 아래서) 조건 지어진 순종으로 끌어내리는 궤변을 늘어놓기 시작했다. 이로부터 종국에는 법칙으로부터의 동기를 능가하는 감성적 충동의 우월성을 행위의 준칙 안에 채용하게 됐고 그리하여 죄를 범하게 되었다.(6:42)

두 인용문 모두에서 등장하는 “의심”과 “궤변”이라는 단어는 칸트가 『종교』를 저술하며 『정초』에서 그가 논했던 성벽에 관한 논의를 떠올리고 있었음을 보여 주고 있다. 그러나 『정초』의 경우 이 의심과 궤

21) 칸트는 경향성과 이성의 이 대립을 “자연적 변증법”이라 명명하고 있다. 4:405 참조.

변에서 법칙에서 이탈하려는 성벽이 바로 도출되는 것처럼 서술하고 있다. 그러나 이 경우 이러한 성벽의 책임은 인간이 아니라 그의 유한성에 물어지게 될 수밖에 없고, 따라서 성벽은 인간이 범한 악이라 할 수 없다. 또한 이러한 성벽에서 기인하는 악의 책임을 인간에게 물을 수 있는 가도 문젯거리로 남게 된다. 이런 구도 하에서는 인간의 도덕적 악을 주제로 논할 수 없다. 반면에 『종교』에서는 의심과 궤변이 인간의 경향성에 의해 자연적으로 나온 것으로 서술되어 있지 않다. 인간은 금지로서 법칙이 주어져 있음에도 불구하고 경향성에 유혹 당해 그것을 의심하고, 궤변을 늘어놓더니 이윽고 스스로 악한 준칙을 채택하는 죄를 범하게 된다. 여기서 칸트는 철저하게 인간의 자발적 선택을 강조하고 있다. 인용문은 악한 준칙의 채택만이 죄인 것처럼 서술하고 있으나 법칙을 의심하고 궤변을 늘어놓은 것 또한 인간이 스스로 범한 것이기 때문에 그 또한 악이라 할 수 있다. 즉, 인간이 악한 것, 인간이 죄를 범한 것은 인간의 유한성 때문이 아니라 전적으로 인간 자신의 선택에 의한 것이다. 물론 “이러한 의사의 엇나감, 다시 말해 악으로의 이러한 성벽의 이성근원은 우리가 탐구할 수 없는 것으로 남는다. [...] 우리에게는 그로부터 도덕적 악이 우리 안에 최초로 나타날 수 있었던 어떤 이해가능한 근거도 없다”(6:43) 『정초』에서와는 달리 악의 근원은 해명 불가능한 것으로 나타난다. 악의 근원을 이해하려는 시도를 포기한 대가로 지켜지는 것은 악의 귀책성이다.

이처럼 성벽을 인간에게 귀책 가능한 것으로 도덕적 악으로 상정함으로써 『정초』에서는 유한한 인간에게 자연 발생적으로 일어나는 것처럼 보였던 타락의 전 과정을 모두 도덕적 악으로 규정하게 된다. 이에 따라 칸트는 악으로의 성벽을 다음 세 가지 타락의 단계로 구분하여 서술한다.

첫째는 “취한 준칙 일반을 준수함에 있어서 인간 심정의 연약함, 또는 인간 자연본성의 허약성”이고, 둘째는 “비도덕적인 동기들과 도덕적 동기들을 뒤섞으려는 성벽, 다시 말해 불순성”이며, 셋째는 “악한 준칙들을 채택하려는 성벽, 다시 말해 인간 자연본성의 또는 인간 심정의 악의

성”이다(6:29).

허약성의 단계에서 인간은 경향성의 유혹을 이기지 못해 준칙을 준수하지 못하게 된다.²²⁾ 그리고 불순성 단계에서는 법칙을 의사 규정의 충분한 동기로 채용하지 않고 자기사랑의 원리를 또한 동기로 채용하여 자신의 경향성을 충족시키는 한에 있어서만 법칙을 따르려 한다. 이러한 단계를 거쳐 이윽고 인간은 의식적으로 도덕법칙을 도덕적이지 않은 다른 동기들 뒤에 놓으려는 의사의 성벽을 가지게 된다. 이것은 “자유로운 의사의 동기들에 대하여 윤리적 순서를 뒤집는 것”이기 때문에 “전도성”이라 할 수 있다(6:30).

성벽에 대한 지금까지의 논의에서 주의를 기울여야 하는 것은 성벽 개념과 마음씨 개념 사이에 명백한 유사성이 있다는 것이다. 두 개념 모두 다음과 같은 성격을 가지고 있다. 먼저, 우리 자유 사용의 제일근거로서 의사가 준칙 안에 어떤 동기를 채택함으로써 결정되는 것이다. 다음으로, 자유 사용의 제일의 근거로서 현실에서 우리가 행하는 모든 행실에 선행하는 것이라는 의미에서 선천적이라 할 수 있다. 마지막으로, 특정한 성벽과 마음씨에서 나온 행위는 모두 동일한 도덕적 성질을 가질 수밖에 없다. 달리 말하면 특정한 성벽과 마음씨는 인간의 행위에 의해 변화할 수 없다는 것이다. 따라서 이 두 개념을 어떻게 구분하고 관련지를 것인가라는 문제가 발생한다. 지금까지 정리한 칸트의 논의가 보여주는 바는 이 두 개념이 대상의 범위에 있어 다르다는 것이다.

칸트의 마음씨 개념에 대해 논하면서 우리는 인간의 마음씨를 직접 관찰할 수 없으므로, 인간의 도덕성에 대한 판단이 확정적으로 내려질 수 없음을 확인하였다. 다음으로 소질 개념은 우리에게 인간이 악하게끔 영향력을 행사하는 그 무엇도 인간 안에 없으며, 인간의 도덕적 악은 인간

22) 우리를 유혹에 굴복케 하는 준칙이 어떤 것일 수 있는지는 생각하기 어렵다. 그 준칙이 ‘유혹의 정도가 강해 그것을 참아내는 것이 극히 고통스러울 경우 고통을 피하고 쾌락을 얻기 위해 도덕법칙을 위배 한다’와 같은 것일 경우, 이 준칙이 악으로의 성벽의 마지막 단계인 심정의 전도성과 어떤 차이를 갖고 있는지 알 수가 없기 때문이다. 아무런 준칙을 갖지 않은 채 우연적으로 일어나는 유혹에 대한 굴복은 칸트의 실천철학과 궤가 맞지 않는 것으로 보인다. 이는 불순성의 경우에도 마찬가지이다. 이 문제에 대해서는 뒤에서 마음씨와 성벽의 관계를 논하며 더 상세히 다룰 것이다.

이 스스로 초래한 것, 그것도 적극적으로 선의 소질을 타락시킴으로써 초래한 것이란 것을 확인시켜 주었다. 마지막으로 성벽 개념이 앞 선 두 개의 개념이 공백으로 남겨 두었던 인간의 보편적 도덕적 성격에 대한 논의를 위해 도입되었다. 그것은 인류가 획득한 도덕적 성격이다. 그러나 성벽 개념과 마음씨 개념의 유사성은 마음씨 개념의 문제를 그 범위만을 달리한 채 동일하게 드러내고 있다. 그것은 우리가 성벽을 어떻게 판정할 수 있는가라는 문제이다. 성벽 개념은 마음씨 개념에서 더 나아가 도덕적 성격에 대한 확정적 판단의 근거가 될 만한 것을 포함하고 있는가? 아래에서 자세히 살펴보게 되겠지만, 이에 대한 대답은 ‘그렇지 않다’이다.

1.4. 악의 보편성에 대한 실패한 증명

칸트는 자연본성적이라는 말에 두 가지 의미로 사용하고 있다. 하나는 우리 자신에 의해 채택된 자유 사용의 궁극적 근거라는 의미이며, 다른 하나는 유적 관점에서 본 인간에게 타당하다는 의미이다. 그러나 이미 지적하였다시피 첫 번째 의미는 도덕의 귀책성 문제와 자연본성적이라는 말의 일상적 의미를 어느 정도 조화시키고 있으나, 두 번째 의미는 그렇게 하고 있지 않다. 자연본성적이라는 말은 어떤 종 전체에게 타당한 성격을 가리킨다. 그러나 어떤 개체가 그 종에 속한다는 이유만으로 특정한 도덕적 성질을 가지게 된다면 과연 그 성질에 대한 책임을 그 개체에게 물을 수 있는가는 의문이다. 그런데 칸트는 도덕적 성질을 다루는 대상이 개별 인간이 아니라 인류여야 함을 전제하고 있다. 그는 자신이 지금 논의하고 있는 것이 인간의 자연본성적인 도덕적 성질이므로, 이에 대해 논의할 때 나오는 결론은 유로서 인간 전체에 대한 하는 것으로 이해되어야 한다고 쓰고 있는데,²³⁾ 이는 자신의 철학적 입장과 자연 본성

23) “우리가 ‘인간은 자연본성상 선하다 또는 악하다.’고 말할 때의 인간은 개별적인 인간으로가 아니라 [...] 오히려 전 인류로 이해되는 것이 적격의 일이고, [...]”(6:25)

적이라는 말의 일반적 의미를 조화시키려는 노력 없이 그저 일반적 의미에만 호소하는 주장이며, 그 논리적 허술함은 명백하다고 할 수 있다. 따라서 인류의 도덕적 성질에 대한 논의가 이루어지기 위해서는 인류의 보편적인 도덕적 성질을 판단하는 것이 과연 가능한 일인지 아닌지, 또한 가능성 여부를 떠나 그에 대해 사고하는 것이 우리에게 과연 필요한 것인지에 대한 설명이 필요하다. 그러나 칸트는 이 문제들에 대해서는 아무런 해답도 내리지 않은 채 인간에게 선이나 악이라는 도덕적 성질을 부여하기 위한 탐구를 시작한다. 그러나 칸트의 탐구는 최소한 표면적으로 보았을 때 명백히 실패하고 있다.

칸트에 따르면 인류의 보편적 도덕적 성질이란 다음과 같은 것이다. 먼저, 그것은 인간에게 책임을 물을 수 있는 것이어야 하기 때문에 필연적이어서는 안 된다. 따라서 도덕적 성질은 인간의 유개념으로부터 추론될 수 없다. 그와 같은 판정은 경험을 통해 내려지는 것이거나 “주관적으로 필연적인 것으로 전제”(6:32)되는 것이다. 또한 이 성질은 도덕적인 한 의사의 준칙들 안에 존립하는 것이며, 이 준칙들은 자유로이 채택된 것이기에 우연적인 것이라 할 수 있다. 그럼에도 불구하고 이 도덕적 성질은 보편적인 것이기에 “모든 준칙의 주관적인 최상의 근거가, 무엇에 의해서든지 간에, 인간성 자체와 함께 짜여 있고, 그리하여 말하자면 그 안에 뿌리박고 있는 것”(6:32)이라 할 수 있다. 이 도덕적 성질은 근본적이라 할 수 있는데, 왜냐하면 그것은 모든 준칙들의 근거이기 때문이며, 또한 자연본성적 성벽으로서 인간의 힘으로는 변화시킬 수 없는 것이기 때문이다. 그리고 그 성질은 악이다. 즉, 인간은 모든 준칙들을 부패시키며 인간의 힘으로는 절멸시킬 수 없는 악성을 가지고 있다는 의미에서 근본적으로 악하다.²⁴⁾ 칸트는 이에 대해 다음과 같은 증명을 제시하고 있다.

무릇 그러한 부패한 성벽이 인간 안에 뿌리박고 있음이 틀림없다는 사실에 관해서는 인간들의 행실에서의 경험이 우리 눈앞에 제시하는 다량의 뚜렷한

24) 그럼에도 불구하고 칸트는 극복 가능성을 완전히 부정하지는 않는다. 인간은 선택해야 하기 때문에 선택될 수 있다.(6:37 참조)

실례들이 있으므로 그에 관한 형식을 갖춘 증명을 우리는 생략할 수 있을 것이다.(6:33)

이 증명이 증명으로서 기능하지 못함은 명백하다. 우리가 경험적 사례를 통해 할 수 있는 최대한의 주장은 악이 세계에 널리 퍼져 있다는 것이지 그것이 보편적이라는 것이 아니다. 게다가 선한 행위와 악한 행위에 대한 칸트 자신의 설명이 경험적 증거에 호소하는 것을 어렵게 한다. 행위 그 자체가 아니라 인간이 행위를 위해 채택한 준칙이 악한 것이다. 그런데 우리는 행위를 관찰할 수 있을 뿐, 준칙을 관찰할 수는 없다. 우리가 할 수 있는 것은 어떤 행위에서 그 밑에 놓여 있는 준칙을 선험적으로 추론해 내는 것뿐이다.²⁵⁾ 그러나 우리는 그 행위에 있어 완벽하게 덕스럽지만 그 최상의 준칙에 있어서는 악한 사람을 생각할 수 있다. 그러므로 우리의 경험은 우리에게 도덕적 성질을 입증할 수 있는 안정적인 증거가 되어 주지 못한다. 칸트의 주장은 우리가 대부분의 사람들에게서 악한 행위를 보고, 소수의 사람들에게서 오직 선한 행위만을 본다고 하더라도 우리는 인류 전체를 악하다고 보아야 한다는 것인데, 이는 선한 행위가 존재함에도 불구하고 그에서 악을 추론해 낼 수 있는 이유에 대한 또 다른 증명을 요구한다. 따라서 경험에 근거한 인류의 보편적 악성에 대한 증명은 불가능하다. 결국 칸트는 마음씨 개념이 갖고 있는 문제가 성벽 개념에서도 해결이 되지 않은 것이다. 하지만 칸트는 여기서 분명히 ‘형식적인 증명’ 또한 가능함을 이야기하고 있다. 즉, 인간 개개인의 도덕적 성격에 대한 판단의 귀납이라는 길을 통하지 않고서도 증명할 수 있는 방안이 있다는 것이다. 칸트가 생략한 이 ‘형식적인 증명’은 어떠한 성격의 것일까? 아래와 같은 두 가지 증명 전략이 가능할 것으로 생각된다.

(1) 칸트의 주장은 우리가 모든 인간 안에서 필연적으로 악으로의 성벽을 발견할 수 있다는 것으로 보인다. 그렇다면 칸트의 주장은 모든 인간

25) “어떤 인간을 악하다고 부르려면, 몇몇의, 아니 단 하나의 의식적인 악한 행위로부터 그 행위의 기초에 놓여 있는 악한 준칙을 선험적으로 추론해내야만 한다.”(6:20)

은 모든 시간과 공간에서 악하다는 선험적 종합 판단이다. 그렇다면 칸트가 『순수이성비판』에서 주장한 바와 같이 이를 정당화하기 위해서는 초월적 논변이 필요하다.²⁶⁾ 그와 같은 증명이 정말 존재하고 칸트가 단지 이를 생략한 것일 뿐이라면 초월적 논변을 재구성함으로써 악의 보편성을 증명할 수 있을 것이다.

(2) 또 다른 증명 전략은 자유와 양립 가능한 악의 근거를 발견해 내는 것이다. 우리는 이 근거가 인간 안에 보편적으로 놓여 있음을 보임으로써 악의 보편성을 증명해 낼 수 있을 것이다. 칸트 자신이 부정적인 방식으로 이 전략을 채택하고 있다. 칸트는 악의 근거에 대한 잘못된 이해를 방지하기 위해 사람들이 생각할 법한 악의 가능한 두 가지 근거를 각각 검토한다. 먼저 칸트는 악의 근거를 인간의 감성과 감성에서 생기는 자연적 경향성들을 검토한다. 그러나 감성과 경향성은 악의 근거가 될 수 없다. 왜냐하면 그것들은 천부적인 것들로 인간이 창시자가 아니기에 인간이 그에 대해 책임을 질 수 없는 것이기 때문이다. 다음으로 칸트는 이성을 검토한다. 도덕적으로-법칙수립적인 이성 자신이 부패해 있다면 그로부터 악이 나올 수도 있기 때문이다. 그러나 이런 일은 절대로 불가능하다. 왜냐하면 “자신이 자유로이 행위하는 존재자라고 생각하면서 또한 그러한 존재자에게 알맞은 법칙(즉 도덕법칙)에 매여 있지 않다고 생각하는 것은, (자유이기 때문에 자연법칙들에 의한 규정은 해당이 없으므로,) 일체의 법칙 없이 작용하는 원인을 생각하는 것이나 마찬가지로 이겠는데, 이런 일은 모순”(6:35)이기 때문이다. 이성이 부패했다는 것은 인간을 도덕법칙에서 풀려난 악마적 존재자로 만들 터인데, 앞서 인격성의 소질에서 보았듯이 도덕법칙에 대한 존경은 인간의 필연적 구성요소인 것이다. 그러나 칸트는 인류의 악성의 가능한 두 설명을 모두 기각한 후 타당한 다른 근거를 제시하지는 않고 있다. 이를 찾아낼 수 있다면 악의 보편적 현존을 보다 설득력 있게 증명할 수 있을 것이다.

26) KrV, A85=B117.

다음 절에서는 이 각각의 전략에 따라 시도된 두 가지 대안적인 ‘형식적 증명’을 소개하고, 두 증명의 한계를 비판할 것이다.

2. 악의 보편성에 대한 두 가지 대안적 증명

개관한 바와 같이 악의 보편성에 대한 칸트의 귀납적 증명은 성공하지 못했기 때문에 대안적인 두 가지 증명 전략이 도출된다. 하나는 칸트의 경험에 대한 호소를 전적으로 부차적인 것으로 보고 칸트를 대신하여 초월적 논변을 제공하는 것이며, 다른 하나의 계열은 칸트의 경험에 대한 호소-넓은 의미에서의 인간학-를 진지하게 받아 들여 악의 보편성에 대한 주장을 보다 타당한 것으로 만들 수 있는 모종의 인간적 특성을 찾아내는 것이다. 전자의 해석을 대표하는 학자는 엘리슨(H. E. Allison)이며 후자의 해석을 대표하는 학자는 우드(A. Wood)이다. 우드가 후자의 해석을 대표할 수 있는 이유는 칸트의 근본악 테제를 칸트의 역사철학 및 사회철학과 직접 연결시킴으로써 논의의 지평을 크게 확장시켰기 때문이다. 아래에서 이 각각의 해석을 대표하는 주석가인 엘리슨과 우드의 해석을 살펴보고, 그 장점과 한계를 논해 볼 것이다.²⁷⁾

27) Wood와는 다른 방식의 인간학적 접근을 시도하는 학자들은 여럿 있으나 이들은 모두 악의 보편성에 대한 증명은 포기한 채 그것이 널리 퍼져 있을 수밖에 없는 이유를 인간의 본성에서 찾아내려 한다. Muchnick(2009)은 인간은 행위에 있어 그 귀결을 고려하지 않을 수 없다는 『종교』의 첫 번째 서론에 나타난 인간학적 특징에서 악의 보편성을 증명하려고 하고 있다. Sussman(2005)은 자유의 사용을 시간 속에서 우리의 성숙(이성을 획득함)과 결부시킴으로써 미성숙한 시기의 우리의 경험이 우리의 악의 근원이라고 주장한다. Grimm(2002)은 감각적 경향성과 이성의 명령에 모두 종속된 인간의 복합적 본성에서 악의 근원을 찾고 있다. 그러나 이 모든 ‘인간학적’ 시도들은 우드만큼 강력한 악의 보편성에 대한 증명을 제공하는데 실패하고 있다. 그러한 인간학적 특징이 있다고 하더라도 여전히 개별 인간들은 다른 도덕적 선택을 내릴 가능성을 갖고 있기 때문이다.

2.1. 악의 보편성에 대한 앨리슨의 초월적 연역

문제에 답하기 위해 앨리슨이 던지는 질문은 다음과 같은 것이다. 왜 칸트는 선을 향한 성벽을 논하지 않았는가? 칸트는 선의 소질과 악으로의 성벽에 대해 논하였는데, 선의 소질이 있다면 그에 대응하는 악의 소질이, 그리고 악으로의 성벽이 있다면 그에 대응하는 선으로의 성벽이 있을 가능성도 있지 않겠는가? 만약 선으로의 성벽의 가능성을 부정할 수 있다면 인간이 보편적으로 악으로의 성벽을 가지고 있음을 증명할 수 있을 것이다.

선의 소질에 대응되는 악의 소질이 칸트의 철학 구도 내에서 나올 수 없다는 것은 명백해 보인다. 칸트의 목적론적인 철학 구도에 따를 때 우리는 자연 또는 신이 인간을 악하도록 창조했다고는 결코 생각할 수 없다. 또한 만약 악을 향한 소질과 같은 것이 있다면 그것은 이성의 사실인 도덕법칙에 대한 존경을 가지고 있지 않은 존재를 상정하는 것일 텐데 그와 같은 ‘악마적 존재자’가 존재할 가능성을 칸트는 분명히 거부한 바 있다.

그렇다면 선으로의 성벽은 어떠한가? 우리가 악을 향한 성벽을 (최소한) 개별적인 인간에게 귀속시킬 수 있는 것이라면, 마찬가지로 선을 향한 성벽을 인간에게 귀속시키는 것도 가능하지 않겠는가? 앨리슨의 증명의 핵심은 인간에게 선을 향한 성벽을 귀속시키는 것은 불가능하다는 것이다. 그의 증명은 다음과 같이 재구성될 수 있다.²⁸⁾

(1) 우리가 가지고 있는 것과 같은 유한한 의지에게 선을 향한 성벽을 귀속시키는 것은 불가능하다. 그와 같은 성벽은 행복의 동기보다 도덕성의 동기를 자발적으로 선호한다는 것인데, 이것은 신성함의 성질이며 이와 같은 신성함을 획득하는 것은 인간의 의지에게는 불가능하다.

(2) 따라서 우리는 인간에게 선을 향한 성벽을 부여할 수 없으며, 도덕

28) 이 재구성은 Muchnick(2009), p. 130을 참조하였다.

적 성질은 선하거나 악하거나 할 수 밖에 없으므로 우리는 인간에게 악을 향한 성벽을 부여해야 한다.

(3) 선을 향한 성벽의 불가능성은 논리적인 것이 아니므로(선을 향한 성벽을 가진 인간을 생각하는 것은 자기 모순적이지 않다), “인간은 자연본성적으로 악하다.”는 주장은 선험적 종합 명제이다.

그러나 이 증명은 성공적이지 못하다. 전제 (1)은 신성함과 선으로의 성벽을 같은 것으로 놓고 있는데 이러한 동일시는 옳지 않기 때문이다. 엘리슨은 (1)을 다음과 같이 설명한다. 인간의 의지는 유한한데, 이는 의지가 감각의 경향성에 영향을 받고 따라서 행복의 추구에 관심을 갖는다는 것을 의미한다. 따라서 감각의 영향을 받는 인간은 “결코 유혹의 가능성과 도덕적 제약의 필요에서 벗어날 수 없다.”²⁹⁾ 이렇게 인간이 신성하지 않다는 것을 증명할 수 있다. 칸트가 『정초』에서 말한 바에 따르면 신성한 의지란 오직 선의 표상에 의해서만 규정될 수 있기에 법칙에 의해 명령받지 않는 의지, 의욕이 법칙과 필연적으로 일치하기 때문에 당위와 관련이 없는 의지이기 때문이다.³⁰⁾ 그러나 이것이 인간은 선을 향한 성벽을 가질 수 없다는 것과 같은 의미는 아니다. 인간은 자신 안에서 도덕적 동기에 저항하는 감성적 동기를 느낌에도 불구하고 감성적 동기에 저항해 자발적으로 도덕법칙을 따를 수 있기 때문이다. 그렇다면 선을 향한 성벽을 가지고 있음과 신성함은 다른 것이다. 머치닉이 지적하는 것처럼 신성한 의지는 선을 향한 성벽과 같은 것일 수 없다. 오히려 신성한 의지는 자신 안에 도덕법칙에 저항하는 아무런 동기도 지니지 않은 의지로서, 그러한 의지에는 어떠한 성벽도 귀속시킬 수 없다.³¹⁾ 그러므로 인간이 신성하지 않다는 것을 증명한다 하더라도 그것이 곧 인간이 악하다는 것에 대한 증명은 되지 않는다.

29) Allison(1990), p. 156

30) “신성한 의지에 대해서는 아무런 명령도 타당하지가 않다. [...] 의욕이 이미 스스로 법칙과 필연적으로 일치해 있기 때문이다.”(4:414)

31) Muchnick(2009), p. 131.

또한 엘리슨의 논변은 『종교』에서 칸트가 제시한 인간의 행위에 대한 설명과 부합하지 않는다. 『종교』에서 칸트는 인간에게 악마적 악이라는 개념을 귀속시키는 것을 거부한다. 악마적인 악은 “도덕법칙을 반역적으로(즉 불복종을 선언하면서) 포기”하는 것(6:36), “악을 악으로서 [그 자체로] 자기의 준칙 안에 동기로 채용하는 마음씨”(6:37)이다. 그리고 이러한 악은 인간이 이성을 가진 자유로운 행위자인 한, 인간에게 귀속시킬 수 없다. 왜냐하면 “이성이 자기 안에 있는 법칙 자신의 권위를 말살하고, 이 법칙에 의한 책무성을 부인할 수 있다는 것인데, 이런 일은 절대로 불가능하기 때문이다. 자신이 자유로이 행위하는 존재자라고 생각하면서 또한 그러한 존재자에게 알맞은 법칙(즉 도덕법칙)에 매여 있지 않다고 생각하는 것은, 일체의 법칙 없이 작용하는 원인을 생각하는 것이나 마찬가지로 이겠는데, 이런 일은 모순이다.” 따라서 악을 향한 성벽은 우리 안에 도덕법칙에 대한 존경이 없다는 것을 의미하는 게 아니라, 우리가 도덕법칙을 자기 사랑의 동기 아래 종속시킨다는 것을 의미할 뿐이다. 인간은 도덕법칙의 명령을 분명히 의식함에도 불구하고 그것을 자기 사랑에 종속시킨다. 마찬가지로 선을 향한 성벽 또한 자기 사랑의 동기가 전혀 존재하지 않음을 의미할 필요는 없다. 인간은 자기 사랑의 동기들의 유혹을 의식함에도 불구하고 도덕법칙에 대한 존경을 동기로 삼아 행동하기를 자발적으로 선택할 수 있다. 엘리슨은 선한 행위는 아무런 감성적 동기의 영향도 받지 않는 행위라고 말함으로써 『종교』에서 제시된 인간의 행위에 대한 설명을 거스르고 있다.

이와 같이 신성함의 의미에 대해서는 『정초』의 논의와 인간 행위의 성격에 대해서는 『종교』의 논의와 충돌한다는 점 외에, Allison의 설명이 지니는 또 다른 문제는 그것이 궁극적으로 인간의 악함의 책임을 인간의 감각적 본성, 유한성에 묻고 있고 따라서 악의 귀책 가능성에 대한 칸트의 논의를 무력하게 만든다는 것이다. 인간이 감성적 존재이기에 신성하지 않고, 그런 의미에서 인간에게 악을 향한 성벽을 귀속시킬 수 있다면 악을 향한 성벽의 궁극적 책임은 인간 자신에게 책임이 물어질 수 없는 것이며, 그것은 도덕적 악이라고 할 수 없는 것이다. 따라서 엘

리슨은 궁극적으로 도덕적 악의 보편성을 초월적으로 연역해 내는 데 실패했다고 볼 수 있다.

엘리슨의 해석의 또 다른 문제는 그의 논의가 칸트의 근본악 테제가 갖는 역사철학 및 사회철학적 함의를 전혀 포착해 내지 못하고 있다는 것이다. 『종교』의 전체적인 주제는 인간의 집합적인 도덕적 발전의 역사에서 종교가 맡는 역할이다. 칸트는 종교가 인간의 도덕적 발전 과정에서 과도적이기는 하지만 매우 중요한 역할을 수행한다고 본다. 그러나 엘리슨의 해석처럼 인간의 도덕적 악이 인간의 유한성에 있는 것이라면 종교가 이를 극복하는데 어떤 긍정적인 역할을 할 수 있는지 알기 어렵다. 그러므로 우리는 칸트 도덕철학의 핵심인 귀책성을 정초하지 못한다는 점과 인간의 도덕적 발전 과정에서 종교를 포함하여 어떤 집합적인 노력이 경주될 수 있는지에 대해 해명하지 못한다는 점에서 엘리슨의 증명을 기각해야 한다.

2.2. 우드의 인간학적 증명

우드에 따르면 Allison의 해석은 신성하지 않음에서 악함을 도출하는데 실패하고 있을 뿐 아니라 인간의 악함을 곧 인간의 유한성과 동일한 것으로 본다는 점에서 칸트의 악에 대한 논의를 진부한 것으로 만들어 버린다는 점에서 문제가 있다.³²⁾ 우드는 엘리슨이 시도한 것과 같은 초월적 논변을 통한 보편적 악의 현존 증명을 칸트가 부정했다고 본다. 칸트는 『종교』에서 악에 관해 논의할 때 초월적 논변이 할 수 있는 역할을 다음과 같이 규정하고 있다.

인간의 자연본성에서 이러한 악으로의 성벽의 현존을 시간상에서 실제적인 인간 의사의 법칙과의 상충에 대한 경험적 증명들을 통해 밝혀낼 수 있다 하더라도, 이 경험적 증명들은 우리에게 이 성벽의 본래적 성질과 이 상충의 근거를 가르쳐주지 못한다. 오히려 이 성벽의 본래적 성질은 자유의사

32) Wood(1999), p. 287 주석 7번 참조.

와 동기로서의 도덕법칙과의 관계에 관련되어 있으므로, 악의 개념으로부터 선험적으로 인식되지 않으면 안 된다.(6:35)

우드는 이에 근거해서 칸트가 악의 현존에 대한 초월적 논변의 가능성을 부정했으며, 악의 현존은 경험적인 증거를 통해 증명될 수 있다고 생각했다는 결론을 이끌어 낸다.³³⁾ 그러나 우드가 염두에 두고 있는 경험적인 증명은 단순한 귀납적 일반화가 아니다. 그는 칸트가 말한 “인간학적 탐구”란 『종교』의 1부에 나와 있는 소질과 성벽에 대한 논의가 아니며, 『종교』에서 칸트가 직접적으로 “인간학적 탐구”를 전개하고 있지 않으므로 그가 여러 저술에서 논한 바 있는 인간의 역사적, 사회적 삶에 대한 논의를 통해 결여된 인간학적 증명을 재구성해야 한다고 주장한다. 그는 『종교』 2부의 다음 구절이 이러한 시도를 정당화해 준다고 본다.

그가 그를 이러한 위험에 빠뜨리고 그 위험에 가두어두는 원인들과 상황들을 살펴본다면, 그가 쉽게 확신할 수 있는바, 그것들은 격리되어서 현존하는 경우의 인간 자신의 자연적인 본성으로부터 오는 것이 아니라, 오히려 인간이 관계를 맺고 교류하는 인간들로부터 오는 것이다. 자연적인 본성의 자극으로는 그의 근원적으로 선한 소질을 그토록 크게 황폐화시키는 본래적 의미의 이른바 걱정들이 인간 안에서 일어나지는 않는다. 그의 필요욕구는 참 작은 것이고, 그러한 필요욕구를 염려하는 인간의 마음상태는 온건하며 평온한 것이다. 인간은 타인들이 그를 가난하다고 여기고 그에 대해 경멸할 것이라고 염려하는 한에서만, 가난하다. 질투, 지배욕, 소유욕 그리고 이것들과 결합되어 있는 적대적인 경향성들은 인간이 다른 인간들 가운데에 있을 때, 그 자체로는 충족한 그의 자연본성을 이내 몰아붙인다. [...] 그들이 인간이라는 사실만으로, 상호 간에 서로 도덕적 소질을 타락시키고 서로를 악하게 만들기에 충분한 것이다. [...] 도덕성의 유지를 목표로 하는 사회를 건설

33) 그러나 칸트의 이 구절은 Wood가 제안한 것과 같은 방식으로 읽을 수 있는지는 의문이다. Wood는 칸트의 문장을 변형하여 “어떤 선험적인 논변도 악의 현존을 증명할 수 없다.”고 인용하고 있다. 그러나 원문은 이와 같은 단정적인 서술이 아니다. 또한 우리는 Allison의 작업을 개념 분석을 통해 존재를 이끌어 내는 시도로 읽을 수 없다. 그러므로 위의 인용문을 통해 Allison의 시도를 부정하는 것은 정당하지 않다고 할 수 있다.

하기 위한 어떤 수단을 찾아내지 못한다면, 개인이 악의 지배에서 벗어나기 위해 제아무리 애쓴다고 해도 악은 그를 부단히 악의 지배 아래로 되돌아갈 위험 중에 붙잡아 둘 것이다.(6:93-94)

사회로부터 격리된 인간이 존재한다면, 그에게 도덕법칙을 위반하는 성향을 귀속시킬 이유가 없다. 따라서 악을 향한 성벽은 인간의 유한하고 감성적인 본성에 의해 설명되는 것이 아니라 인간의 사회적인 성격에 대한 칸트의 인간학적 주장에 의해 설명된다. 인간의 “다른 사람의 눈에 무가치한 존재로 보일 수 있다는 공포”가 그로 하여금 “사회적 비교와 항쟁”을 불러일으키게 한다.³⁴⁾ 이로써 인간은 “타인들이 (자신의 위에서는) 우월성을 얻으려 애쓸 것이라는 걱정[...]으로부터 점차 자기가 타인 위에 서는 우월성을 획득하려는 부당한 욕구”(6:27)를 얻게 된다. 그러나 인간이 자유로운 이성적인 존재인 한 그의 가치란 다른 사람과의 비교가 아니라 도덕법칙에 따라서만 평가될 수 있는 것이다. (5:76) 그런데 사회에서 인간은 타인과의 비교를 통해 자신에게 가치를 부여하게 됨으로써 자신을 이성적인 존재자로서 올바르게 취급하지 않게 되어 도덕법칙이 아니라 자기사랑의 원리를 따르게 된다.³⁵⁾ 이것이 바로 악의 근원이다. 따라서 우드에 따르면 근본악은 사회적 현상으로서, 인간이 사회적 상황에 처해 있다는 바로 그 사실에서 생겨나는 것이다. 따라서 우드는 근본악이 사회적 “성격”과 “근원”을 가지고 있다고 주장하면서,³⁶⁾ 이 근원을 칸트가 『세계 시민적 관점에서 본 보편사의 이념』(이하 『이념』)에서 제시한 “비사회적인 사교성(ungesellige Geselligkeit)”과 동일시한다.

『이념』에서 칸트는 자신의 목적론적 자연관에 근거해 인간의 역사를

34) Wood(1999), p. 288.

35) 악의 근원을 타인과의 항쟁에서 찾는 우드에게 가해질 수 있는 비판 중 하나는 우드가 인간이 타인이 아니라 자기 자신에게 행하는 악(ex. 자살)을 설명하지 못한다는 것이다. 그러나 우드의 논의의 핵심은 인간이 타인과의 비교의 과정에서 자신 및 타인을 이성적 존재자로서 올바르게 대하지 못하게 되는 것이 악의 근원이라는 것이다. 인간이 이성적 존재자로서 법칙에 따르지 않고 자기사랑의 원리에 따르게 될 때 인간은 지금의 고통에서 벗어나기 위해 자살을 택할 수 있게 된다. Wood(2009), p. 167.

36) Wood(1999), p. 289.

다음과 같이 설명하고 있다. 자연은 생명체의 모든 자연적 소질이 언젠가는 완전하게, 목적에 맞게 발현되도록 결정되어 있다. 그런데 인간의 경우 이 자연적 소질은 개인(Individuum)이 아니라 유(Gattung) 속에서만 완전하게 계발된다. 왜냐하면 인간의 이성은 그 완전한 계발을 위해 개인의 수명을 훨씬 능가하는 오랜 시간을 필요로 하는데, 만약 이 소질의 유적 계발 가능성을 부정한다면 이 소질은 인간에게는 쓸모가 없는 것이며, 따라서 목적론적인 이해에 위배되기 때문이다. 이렇게 하여 유적으로 완전히 계발된 인간의 자연적 소질은 사회의 합법칙적인 질서라는 형태를 갖게 될 것인데, 이를 위해 자연이 사용하는 수단은 인간들 상호 간의 항쟁(Antagonism)이다. 인간은 자신을 사회화하려는 경향을 갖고 있으며, 사회적 상태 속에서 인간은 자신의 자연적 소질을 계발하려고 한다. 그러나 인간은 동시에 자신의 의도대로만 행동하려는 반사회적인 특성도 갖고 있다. 따라서 인간은 각자 타인의 저항에 부딪치게 되고, 이 저항에 의해 나태해지지 않고 명예욕, 지배욕, 소유욕 등에 의해서로 대립하고 불화함으로써 그 결과 각종 진보를 이루게 되며, 궁극적으로 인간의 목적인 도덕적인 공동체의 건설로 나아가게 된다. 결국 인간의 반사회적인 성격이 도덕적인 사회를 건설하게 되는 것이므로 이를 “비사교적 사교성”이라고 부를 수 있는 것이다.³⁷⁾

악의 근원과 성격을 비사교적 사교성에 연결시킴으로써 우드는 악의 보편성을 보다 잘 설명할 수 있게 된다. 악의 보편성은 인간이 사회에서 소통하는 존재자라면 갖게 되는 성격이다. 우리는 사회에서 고립되어 살아가는 인간을 상상할 수는 있지만, 우리가 알고 있으며, 알게 될 가능성이 있는 어떠한 인간도 사회에 속한 인간이기 때문에-우리가 어떤 인간을 알게 되고 그와 교류하게 되는 순간, 그는 이미 사회에 속한 것이다-, 우리는 주관적으로 인간악의 보편성을 상정할 수 있게 되는 것이다.

또한 이를 통해 우드는 인간이 도덕법칙의 명령을 위반하는 이유를 일관되게 설명해 낸다. 도덕법칙이 명하는 것은 궁극적으로 우리 자신과

37) 『이념』, 19-22

타인을 인간이 이성적 존재자이기에 갖는 궁극적인 가치에 따라 취급하라는 것이다. 그런데 비사교적 사교성으로 인해 타인의 시선에 의해 자기 자신의 가치를 규정하게 된 인간은 자기 자신을 타인보다 우위에 둘으로써 자신의 가치를 얻으려고 한다. 이것이 인간이 다른 사람은 따르기를 원하는 규칙에서 자신을 예외로 만드는 이유이며, 타인을 목적으로 대하지 않고 단순한 수단으로 대하는 이유이며, 타인의 목적과 조화를 이룰 수 없는 목적을 설정하는 이유이다.³⁸⁾ 이렇게 인간은 자신을 예외적인 존재로 취급함으로써 자신의 경향성을 도덕법칙을 거슬러 충족시키려 하는 것이다.

이처럼 우드의 해석은 악의 보편성을 잘 설명해 내고 있다는 것과 악의 사회적 성격과 칸트 윤리학의 기본적인 명제들을 조화시키고 있다는 점에서 장점을 가지고 있다.

그러나 우드는 다음 질문에 답을 해야만 한다. 이러한 해석은 악의 책임을 신이나 자연, 또는 사회에 귀속시키는 것이 아닌가? 만약 선한 사회에서만 악이 없어질 수 있다는 이야기는 인간이 선한 사회에서 태어나면 그는 악해지지 않는다는 것을 의미한다. 그렇다면 인간이 악한 이유는 궁극적으로 사회가 선하지 않기 때문인가? 우드도 이 문제를 인식하고 있다. 그는 자신의 주장이 칸트의 개인주의적 도덕관을 부정하는 것은 아니라고 주장한다. “악의 사회적 성격과 근원은 [...] 우리에게 자연본성, 섭리 또는 타인들에게 우리의 행위에 대한 책임을 물을 여지를 주는 것이 아니다.”³⁹⁾ 그와 같은 시도는 “우리가 자연 목적들을 (신이나 자연과 같은) 자발적인 동력인(voluntary efficient causes)의 의도된 효과로 이해하려고 한다면, 우리의 반성적 판단을 이끄는 자연 목적론의 규제적 역할을 전적으로 오해하는 것이 될 것이다.”⁴⁰⁾ 우리는 자연, 신 또는 사회를 의지를 가진 것으로서 지성적으로 인지할 수 없기에 마찬가지로 우리는 이들에게 어떤 책임을 부여할 수도 없는 것이다. 그러나 이처럼 칸트의 도덕적 개인주의를 재단언한다고 해서 우드의 논

38) Wood(1999), p. 288.

39) Wood(1999), p. 289.

40) Wood(1999), P. 289.

변이 악의 책임을 사회나 섭리에 귀속시키고 있다는 의심이 사라지는 것은 아니다.

아래에서 우드 증명의 문제점을 두 가지로 나누어 상세히 논할 것이다. 우선, 과연 자연 상태의 인간이 선하다고 할 수 있는지가 문제가 된다. 우드의 논변은 기본적으로 자연 상태의 인간이 선하다는 가정에 기대고 있는데, 칸트의 다른 많은 구절들이 이러한 우드의 가정을 반박하고 있기 때문이다. 다음으로, 설령 우드의 논변이 충분한 문헌학적 증거를 가지고 있다고 하더라도 위에서 제시된 악의 귀책성 문제를 해명하지 못한다면 그것은 칸트의 근본악 테제에 대한 철학적 정당화가 될 수 없을 것이다.

먼저, 자연 상태의 인간을 과연 선하다고 볼 수 있는가라는 문제부터 논해 보자.

우드의 논변을 떠받치는 전제는 인간이 전사회적 삶 또는 자연 상태에서 선하였다는 생각이다. 그러나 자연 상태에서의 인간의 “온건하며 평온한” 상태를 과연 ‘선’한 상태라 할 수 있을까? 칸트는 『도덕철학 강의(Vorlesungen zur Moralphilosophie)』에서 자연 상태의 인간에 대해 다음과 같은 예를 들고 있다.⁴¹⁾ 여기서 칸트는 다음과 같은 예를 든다.

행복 또는 불행, 만족 또는 불만족의 가장 큰 원천은 다른 사람과의 관계에 놓여 있다. 만약 마을의 다른 모든 사람이 상한 치즈를 먹는다면 나도 그것을 만족스럽고 즐거운 마음으로 먹는다. 그런데 다른 사람은 잘 먹지만 나만 그러한 유감스런 상황에 있다면 나는 이를 불행이라 여길 것이다. 따라서 모든 행복과 불행은 우리 자신에 달려 있으며, 우리의 마음이 상황을 받아들이는 방식에 달려 있다.(27:367)

우드의 주장대로 자연 상태에서 만족한 인간이 선하다고 할 수 있다면, 상한 치즈를 만족스럽게 먹는 사람은 도덕적으로 선한 사람이다. 그러나 우리는 상한 치즈를 만족스럽게 먹는 것과 도덕법칙에 대한 존경에서 행

41) 이 예는 Stephen Grimm을 참조하였다.

위하는 것이 동일한 도덕적 가치를 갖는다고 볼 수 없다. 상한 치즈를 먹는 사람의 예에서 여하한 도덕적 선을 찾아내는 것을 어렵다. 도덕적 마음씨의 힘이 증명될 수 있는 것은 오히려 경향성이 도덕법칙에 반해 행위할 것을 요구할 때이다. (6:35) 자연상태의 인간은 도덕법칙에 거스르는 행위를 하지 않을 수도 있으나 그것이 도덕법칙에 대한 존경에서 나온 행위임을 증명할 수는 없다. 상한 치즈를 아무런 불평 없이 먹게 하는 행위 준칙은 도덕법칙을 따르는 준칙일 수도 자기사랑의 원리를 따르는 준칙일 수도 있다.

그런데 만약 자연 상태에 존재하는 인간이 선하고 따라서 그의 의사의 기본 준칙 또한 선하다면 대체 왜 그가 타인과 접촉하게 되었을 때 악하게 된단 말인가? 칸트가 말한 바처럼 최상의 준칙이 악하기에 우리가 악으로의 성벽을 근절시킬 수 없다면⁴²⁾, 최상의 준칙이 선한 인간이 선하기를 그만두고 악해지는 것 또한 있을 수 없는 일이다.

따라서 인간이 사회를 형성했을 때 법칙에 반하는 행위를 하게 된다면 인간은 자연 상태에서부터 이미 악한 준칙, 감각적 경향성을 도덕법칙보다 우위에 두는 준칙에 따라 행위했다고 가정하는 것이 더 타당할 것이다. 그가 반법칙적인 행위를 하지 않은 것은 전사회적 삶이 그런 행위를 필요로 하지 않았기 때문이지 그가 선하기 때문이 아니다. 그렇다면 자연 상태에서의 인간은 선한게 아니라 오히려 악하며, “비사교적 사교성” 또한 악의 근원이 아니라 자연 상태에서 인간이 이미 갖추고 있던 악성이 인간이 사회적 삶을 살게 됨에 따라 이전에는 없던 형태로 나타난 것이라 할 수 있다.

다음으로, 우드의 논변이 악의 책임을 인간에게 지우는데 실패하였고, 따라서 궁극적으로 악이 보편성을 증명하는데 실패했음을 살펴보자.

우드는 자신의 논변이 악의 책임을 개인에게 돌리지 않고 사회에 돌리고 있다는 주장에 대해 사회적 조건이 악으로의 근본적 성벽을 발달시키는 데 있어 필연적인 맥락을 제공한다는 것과 사회가 우리로 하여금 악한 준칙을 채택하도록 강제한다는 주장은 다른 것이라고 반박한다.⁴³⁾

42) “(이 성벽이 근절되기 위해서는) 최상의 준칙이 선의 준칙이어야만 할 것인데, 그 준칙이 저 성벽 자신에서 악한 것으로 상정되고 있으니 말이다.”(6:31)

타인의 존재는 악한 준칙의 선택에서 “필연적인 역할”을 담당하지만 그 사실이 우리의 잘못된 선택에 대한 면죄부를 주지는 않는다는 것이다. “근본악은 [...] 우리가 사회적 존재인 한 우리에게 귀속된다.”⁴⁴⁾ 그러나 어떤 것이 무언가의 형성에 필연적인 역할을 하지만 그것을 결정하지는 않는다는 것은 대체 무슨 의미일까? 자연 상태에서 선하던 인간은 일단 사회적 삶을 살게 되면 타인과의 비교에서 자신의 가치를 찾는 경향성을 발전시키게 되고 이로 인해 타인을 목적으로 대하는 악을 범하기에 이른다. 그렇다면 사회는 한 번 접하면 그에 대한 근절될 수 없는 경향성을 발생시키는 마약과 같은 작용을 하는 것이라 할 수 있을까? 그러나 인간과 사회의 관계는 인간과 마약의 관계와는 분명히 다르다.

다음과 같은 예를 구성해 비교해 보자. 마약이 무엇인지, 그것이 얼마나 위험한지 전혀 모르는 인간이 있다고 하자. 그는 우연히 마약을 접하고, 마약의 쾌락을 느끼게 된다. 그것이 초래할 위험을 모른 채 그는 마약을 계속해서 복용하게 되고, 결국 그는 마약 중독에 빠져 자신의 의무를 수행할 수 없는 인간이 되어 버리고 말았다. 이 경우 의무를 수행하지 않는 그의 잘못의 책임은 어디에 있는가? 마약인가, 복용자인 그 자신인가? 이 경우에도 물론, 책임은 마약이 아니라 마약을 복용한 그 행위자에게 있다. 자신의 의무에 배치되지 않는 한 쾌락을 추구하는 것은 죄가 아니다. 게다가 그는 자신의 의무를 위반할 것을 의도하지도 않았다. 그는 마약이 그에게 어떤 영향을 끼칠지 몰랐기 때문이다. 또한 마약의 중독성은 너무나 강력해서 우리는 마약 중독자가 자신의 의지로 마약의 힘을 이겨내었어야만 한다고 생각할 수 없다. 그러나 이 모든 맥락들도 그의 책임을 면해 주지 않는다. 그에게는 처음에 마약을 복용한 후 아직 중독에 이르기 전에 비록 쾌락은 주지만 어떤 부작용이 있을지 알 수 없는 물질을 멀리하기로 결정할 수도 있었기 때문이다. 그는 낯선 물질을 복용하는 것에 대해 충분한 주의를 기울이지 않았다. 바로 이 사려의 부족이 그를 죄인으로 만든다.

사회를 위 예에서의 마약과 같은 것으로 이해할 수 있을까? 그러려면

43) Wood(2009), pp. 168-169.

44) Wood(1999), p. 287.

사회를 인간이 그 장점을 발견했음에도 불구하고, 그 미지의 위험성을 고려해 즉시 해체해야 마땅한 것으로 이해해야 한다. 그러나 사회에 대한 이러한 이해는 칸트의 생각과는 분명히 배치되는 것이다. 선의 근원적 소질에서 칸트가 이야기하고 있듯이 인간의 사회성은 분명 여러 가지 문제를 발생시키기는 하나 궁극적으로 인간의 발전을 가능하게 하는 것으로써 타락만이 귀결되는 마약과는 다른 것이다. 비록 문제가 발생하더라도 장기적으로 더 좋은 선을 가능하게 하는 것이라면, 인간이 사려를 기울여 사회를 피해야 할 이유란 존재하지 않는다. 따라서 우리는 마약을 복용하는 것과는 달리 사회적 삶을 살아가는 것에 대해 사려의 부족을 이유로 인간에게 도덕적 책임을 지울 수 없다. 그리고 사회적 삶이 발생시키는 도덕적 타락 또한 인간에게 책임이 있는 것이 아니다.

이처럼 사회와 도덕적 책임의 관계를 마약의 예를 통해 이해할 수 없다면, 우리는 이를 어떻게 이해해야 하는가? 우드는 이에 대해 어떠한 분명한 설명도 제시하지 않는다. 이처럼 사회가 정확히 어떻게 기능하는지에 대한 분명한 규정이 없기 때문에 우드의 주장은 악의 책임을 사회로 돌리고 있다는 비판을 피하기 어렵다.⁴⁵⁾

이 밖에도 우드의 주장이 설득력을 얻으려면 전사회적 삶에서 사회적 삶으로의 이행 과정에 대한 보다 상세한 재구성이 필요하다. 인간의 도덕적 성격은 모든 자유 사용에 선행하는 궁극적인 근거에 따라 결정된다. 이것은 인간의 도덕적 성격이 쉽게 변하지 않는다는 것을 의미한다. 악에서 선으로의 이행, 그리고 선에서 악으로의 이행은 혁명적인 사태이다. 사회로의 진입과 더불어 인간 안에 남과 비교하는 경향성이 싹튼다는 우드의 설명은 이 혁명의 성격을 기술하기에는 많이 부족하다. 경향성이 싹이 트는데서 나아가 어떻게 이것이 단 번에 인간의 심성을 뒤집어엎는지를 보여주어야 하는 것이다.

이처럼 많은 문제점을 가지고 있음에도 불구하고 우드의 해석은 악을 인간의 사회성과 관련지어 설명함으로써 악의 보편성에 대해 앨리슨보다는 설득력 있는 증명을 제공하고 있다. 사회와 아무런 접촉도 없이 살아

45) Grenberg(2009), p. 176.

온 인간이 존재할 가능성은 매우 적으며 설령 그러한 인간이 존재했다 하더라도 우리와 접촉하게 되는 순간 그에게서 악이 나타날 것이므로, 우리는 최소한 우리가 알고 있으며, 알게 될 가능성이 있는 모든 인간이 악하다고 주장할 수 있다. 또한 우드의 주장은 칸트의 역사철학 및 사회철학과 칸트의 도덕적 행위 이론을 연결시킬 수 있는 풍부한 논점을 제공해 준다. 인간의 행위는 늘 특정한 사회적 맥락에서 발생함을 생각해 보면 우드의 접근 방식을 통해 칸트의 실천철학의 내용을 더욱 풍부하게 할 수 있을 것이다.⁴⁶⁾

그러나 우드의 주장은 칸트의 다른 여러 주장들과 상충되는 지점이 있으며, 그 중에서도 특히 귀책성의 문제는 도덕성에 대한 칸트의 가장 근본적인 생각과 충돌하는 것으로서 우드의 주장을 받아들이기 어렵게 한다.

3. 악의 보편성에 대한 또 다른 증명

앨리슨과 우드의 해석은 모두 ‘어떻게 악의 책임을 인간에게 물을 것인가?’라는 물음에 답하는 데 실패하고 있다. 앨리슨은 악의 근원을 인간의 유한성에서 찾고 있으며, 우드는 인간의 사회적 삶에서 찾고 있다. 그러나 이 두 입장은 모두 악의 책임이 궁극적으로 우리 자신에게 물어져야 한다는 칸트의 생각과 충돌하고 있다. 따라서 악의 책임이 우리들 자신에게 있다는 사실을 분명히 견지하면서도 악의 보편성에 대한 칸트의 주장 또한 설득력 있게 뒷받침할 수 있는 대안적인 해석을 구성해야 한다. 이를 위한 실마리가 될 수 있는 것은 위에서 언급한 바 있는 ‘악한 마음씨’와 ‘악으로의 성벽’의 유사성 문제이다. ‘악한 마음씨’는 인간이 선하거나 악하다는 것이 무슨 의미인가를 분명히 하고 있으나 어떻

46) 특히 Rossi가 『종교』 3부에서 칸트가 제시한 바 있는 윤리적 공동체를 현실적인 사회제도로 이해하려고 시도한 것은 칸트의 실천철학을 구체적인 사회이론으로 연결시키려는 주목할 만한 노력이라 할 수 있을 것이다. Rossi(2005) pp. 71-85 참조.

게 우리가 어떤 인간의 도덕적 성격을 판단할 수 있는가라는 물음에는 답하지 않았다. 마찬가지로 ‘악으로의 성벽’은 인류가 악하다는 것의 의미가 무엇인지를 분명히 하고 있으나 우리가 어째서 인류가 악하다고 판단할 수 있는가라는 물음에는 답하는데 실패하였다. 이처럼 두 개념은 논의의 범위에서는 다르지만 그 성격에 있어서는 동일한 문제를 가지고 있는 것이다. 칸트 자신은 마음씨와 성벽이 어떻게 구별되는지에 대해 논하지 않았다. 개별 인간의 도덕성을 논하는 맥락에서 마음씨 개념이 사용되고 있으며, 인류 전체의 도덕성을 논하는 맥락에서 성벽 개념이 사용되고 있을 따름이다. 하지만 과연 이런 맥락의 차이를 곧바로 두 개념의 의미의 차이로 연결시켜, 이 두 개념이 논의 대상의 범위에서만 차이를 갖는 것이라고 단정할 수 있는가?⁴⁷⁾ 악으로의 성벽에 대한 칸트의 서술을 면밀히 살펴보면 논의 대상의 범위 외에도 악한 마음씨로부터 악으로의 성벽을 구별할 실마리가 존재한다. 이 실마리를 통해 악으로의 성벽을 새롭게 이해한다면, 악한 마음씨 개념이 가진 도덕성 판정 불가능성 문제를 해소하고 인간 악의 보편성에 대한 새로운 증명을 제시할 수 있다.

칸트는 ‘악한 마음씨’를 우리 의사의 준칙 안에서 도덕적 동기를 자기 사랑의 동기에 종속시키는 것이라 정의한다. 그리고 ‘악한 마음씨’는 인간의 모든 자유 사용의 궁극적 근거로서 더 이상 그 근거가 물어질 수 없는 것이다. 그런데 칸트는 ‘악으로의 성벽’에 대해 논하면서 준칙 안에서 도덕적 동기를 자기 사랑의 동기에 종속시키는 것을 또한 ‘심정의 전도성’이라 칭하고 있다. 즉, ‘악한 마음씨’와 ‘심정의 전도성’은 모두 인간이 동기들의 올바른 관계가 역전된 잘못된 의사의 준칙을 택한다는 동일한 사태를 가리키고 있다. 그런데 문제는 이 전도성이 ‘심정의 허약성’과 ‘심정의 불순성’이라는 악으로의 성벽의 귀결로 다뤄지고 있다는 것이다. ‘허약성’과 ‘불순성’이 ‘전도성’의 근거이다. 그러나 칸트는 ‘악한 마음씨’에 대해서 그에 대해 더 이상의 근거를 물을 수 없다고 하

47) Muchnik(2009)은 이 두 개념의 차이를 논의 범위의 차이로 환원시킨 후 이에 기반해 자신의 논지를 전개하였으나 아래에 서술하였듯이 두 개념 사이에는 논의 범위로 환원되지 않는 차이들이 존재한다.

였다. 그렇다면 우리는 서술이 가리키는 유사성에도 불구하고 두 개념이 다른 두 사태를 가리키고 있다고 보아야 하며, 이 두 사태가 각각 무엇인지, 그리고 ‘악한 마음씨’와 ‘악으로의 성벽’이 어떤 관계인지를 분명히 해야 한다.⁴⁸⁾ 개별 인간의

분석의 과제는 다음 두 가지이다. ‘악한 마음씨’와 ‘악으로의 성벽’이 논의 대상의 차원뿐 아니라 어떤 점에서 다른지를 밝히는 것, 그리고 이 차이를 확인한 후에 개별 인간의 ‘악한 마음씨’가 인간 종의 ‘악으로의 성벽’으로 어떻게 연결되는지를 보이는 것이다.

3.1. ‘악한 마음씨’와 ‘악으로의 성벽’

먼저 다음과 같은 질문을 던져 보자. ‘심정의 허약성’은 과연 도덕적 악이라고 할 수 있는가? 칸트는 도덕법칙을 동기로써 채택하되 그에 따라 살지 못하는 심정의 허약성을 인간에게 책임이 있는 도덕적인 악으로 보고 있다. 도덕법칙을 동기로 채택하였다는 말은 준칙 안에서 도덕법칙의 동기를 자기사랑의 동기보다 우위에 두었다는 것을 의미한다. 그러나 이는 칸트가 정의한 ‘선한 마음씨’와 같은 것이다. 그렇다면 허약성은 전혀 도덕적으로 악한 것이 아니며 허약성으로 인해 저지른 악은 우리에게 책임을 물을 수 없는 것, 즉, 악한 행위가 아닐 것이다. 이것은 우리의 도덕적 직관에도 잘 부합하는 사고방식이다. 우리는 선한 의도를 가지고 있었으나 의지의 허약함으로 인해 이를 그 의도대로 행하지 못한 사람을 의도에서도 이미 악한 사람보다 덜 악하다고 판정한다. 이처럼 의지의 허약함은 우리의 도덕적 책임을 덜어주는 것으로 상정된다. 그러나 칸트는 ‘심정의 허약성’을 분명히 인간이 스스로 범한 도덕적 악으로 규정하고 있다. 도덕적 악이란 도덕법칙이 아닌 자기사랑을 동기로 채택

48) 칸트가 근본악을 악으로의 성벽과 동일시하고 있음에도 불구하고 우드는 악으로의 성벽 개념에 주목하지 않는다. 그는 ‘비사교적 사교성’을 악으로의 성벽 개념과 연결시켜 풀이하려는 시도를 하고 있지 않은데, 이것은 우드 해석의 설득력을 떨어뜨리는 또 다른 문제점이다.

하는 것이다. 그런데 이는 ‘심정의 허약성’에 대한 칸트의 정의와 반대되는 것이다.

‘심정의 불순성’에 대해서도 마찬가지로 문제가 제기된다. ‘심정의 불순성’은 도덕법칙을 순전한 동기로 삼지 못하고 도덕법칙을 준수하는데 있어 자기사랑의 동기 또한 구하는 것이다. 그런데 유한한 존재인 인간은 결코 도덕법칙만을 자기 자신의 준칙에 동기로서 채택할 수 없다. 인간은 또한 자기사랑의 동기 또한 준칙 안에 채택해야 하며, 문제가 되는 것은 이 동기들의 종속 관계이지 이 두 동기가 준칙 안에 동거하고 있다는 것이 아니다. 동기들의 윤리적 순서가 뒤집히는 것은 ‘심정의 전도성’에 이르러서이기 때문에 ‘불순성’의 단계에서는 아직 도덕법칙의 동기가 자기사랑의 동기보다 우위에 있다고 보는 것이 타당하다. 그렇다면 다시금 ‘불순성’이 어떤 의미에서 윤리적 악일 수 있는지가 문제가 된다.

또 다른 문제는 세 가지 성벽들이 어떻게 단계를 이루며 상승해 가는지 가 불분명하다는 것이다. 칸트는 마치 자기사랑의 동기가 점점 그 힘을 키워 이우고 도덕법칙의 동기를 압도하는 것처럼 서술하고 있다. 허약성의 단계에서 자기사랑은 아직 도덕법칙을 압도하지 못한 채 이따금 그에 저항을 시도할 뿐이다. 불순성의 단계에서 자기사랑은 도덕법칙과 어깨를 나란히 하게 된다. 그리고 전도성에서 이우고 자기사랑은 도덕법칙보다 더 큰 힘을 인간에게 행사하게 된다. 그러나 이는 선과 악에 대해 칸트가 가지고 있는 기본적인 생각과 부합하지 않는다. 허약성과 불순성의 단계에서 아직 인간은 도덕법칙을 자신의 동기로 채택하고 있다. 그것은 곧 도덕법칙과 자기사랑이 올바른 순서를 이루고 있다는 것이다. 전도성의 단계에서는 이 순서가 뒤집힌다. 그런데 이것은 단계적인 심화의 과정이라기보다는 단 한 번의 극적인 변화라 할 수 있다. 따라서 허약성과 불순성은 전도성으로 이어지는 단계로 볼 수 없다.

이처럼 악한 마음씨와 악으로의 성벽 사이에 존재하는 많은 충돌들은 이 두 개념을 개별 인간과 인류의 도덕성이라는 분석의 층위만 다를 뿐, 그 타당성에 있어서 동일한 위상을 갖는 개념으로 이해하는 것을 어렵게

한다.⁴⁹⁾ 따라서 악한 마음씨와 악으로의 성벽은 분석의 층위에서 다른 것이 아니라 그 타당성에서 차이가 나는 개념으로 이해해야 한다. 즉, 둘 중 어느 한 개념이 다른 개념보다 더 도덕에 관해 설득력 있는 설명을 제공하고 있으며 다른 하나는 이 설명을 전제로 해서만 의미를 갖는 것으로 이해되어야 한다.

이러한 해석의 실마리를 우리는 우드가 악의 근원을 인간의 사회적 삶 속에서 찾으며 인용한 『종교』의 한 대목에서 발견할 수 있다.

그가 그를 이러한 위험에 빠뜨리고 그 위험에 가두어두는 원인들과 상황들을 살펴본다면, 그가 쉽게 확신할 수 있는바, 그것들은 격리되어서 현존하는 경우의 인간 자신의 자연적인 본성으로부터 오는 것이 아니라, 오히려 인간이 관계를 맺고 교류하는 인간들로부터 오는 것이다. [...] 그들이 인간이라는 사실만으로도, 상호 간에 서로 도덕적 소질을 타락시키고 서로를 악하게 만들기에 충분한 것이다.(6:93)

우드는 이 구절에서 인간의 상호 교류가 인간을 악하게 만든다는 주장의 근거를 찾고 있다. 그러나 이 구절은 다른 식으로 해석될 수도 있다.⁵⁰⁾ 다른 인간들이 인간을 악하게 만든다는 것은 다른 아니라 “위험”에 빠져 있는 인간이 자신이 왜 그러한 상황에 빠져 있는가에 대해 그 자신이 “확신”하게 된 것이다. 즉, 이 구절에서 칸트가 타인의 영향에 대해서 말하고 있는 바는 칸트 자신의 생각이 아니라 이미 “악한 원리의 지배”를 받고 있는 인간이 자기 자신의 상태를 이해하기 위해 스스로에게 확신시킨 생각일 수도 있는 것이다. 그는 자신이 악한 원리의 지배를 받고 있는 이 상황을 그 자신이 초래한 것이 아니라 그가 교류하고 있는 다른 인간들에게서 온다고 생각한다. 칸트는 여기서 “자신의 위험천만한 도덕적 상태의 원인이 자기 자신이 아니라 믿게끔 자기 자신을 속이는 도덕적 행위자의 사고 과정을 묘사하고 있는 것이다.”⁵¹⁾ 이러한 해석의

49) 이 지점에서 나는 Muchnik의 분석에 동의하지 않는다. Muchnik은 각각 마음씨를 ‘선’에 악으로의 성벽을 ‘최고선’에 연결시키면서 두 개념을 도덕성의 각각 다른 두 측면을 다루고 있다고 지적하고 있으나 이 주장은 악한 마음씨와 악으로의 성벽 사이에 존재하는 모순점들을 전혀 다루지 않고 제기되었다는 점에서 타당성을 갖지 않는다.

50) Grenberg(2009), pp. 177-178.

타당성은 칸트가 위 문장을 쓰기 바로 직전에 “인간이 이렇게 위험천만한 상태에 있는 것은 그 자신의 탓”(6:93)임을 분명히 하고 있다는 데서 더 강화된다. 그렇다면 우드와는 반대로 “그들이 인간이라는 사실만으로, 상호 간에 서로 도덕적 소질을 타락시키고 서로를 악하게 만들기에 충분한 것”이란 구절을 우리는 이미 도덕적으로 악한 인간이 자신의 책임에서 벗어나기 위해 내린 결론으로 이해할 수 있다. 사회 상태에 있는 악한 인간은 자신을 기만하여 자기 자신에게 책임이 있는 악을 타인의 탓으로 돌린다. 칸트가 도덕적 악의 핵심적 성격을 기만성에서 찾고 있다는 것은 이러한 결론을 뒷받침한다.

이 선천적 죄과는 인간 심정의 모종의 간악성(惡한 犯罪)을 그 성격으로 갖는다. 즉 자기 자신의 선한 또는 악한 마음씨들로 인해 자기 자신을 기만하며, 그의 행위들이 그 준칙들에 따라서 초래할 수도 있었던 악을 결과로 갖지만 않으면, 자기의 마음씨로 인해 불안해하기보다는, 오히려 법칙 앞에서 자신을 정당하다고 여기는 간악성을 그 성격으로 갖는다. (6:38)

이러한 사람들은 사실은 악한 준칙을 채택한 악인들이지만 우연히 “무능력, 기질, 교육, 유혹으로 이끄는 시간과 장소의 상황이 (즉 우리에게 귀책될 수 없는 것들이) 그들을 패악으로부터 멀리 지켜주지 않았다면, [...] 사유방식[성향]에 따라 똑같은 패악들이 그들에 의해서도 저질러지지 않았을까에 대해서는 규명”해 보지 않는다(6:38). 여기서 ‘악으로의 성향’의 첫 번째 단계인 ‘유혹’에 대한 언급이 등장하고 있음에 주목해야 한다. 실제로는 악하지만 유혹에 직면하지 않아 아직까지 악한 행위를 한 번도 하지 않은 인간은 자기 자신을 선하다고 기만한다. 그가 유혹에 직면해 악한 행위를 하게 됐을 때 그는 자신에게 그 행위는 유혹에 의한 것이라고 변명하려 할 것이다. 그러나 그가 유혹에 굴복했다는 사실이 이미 그 안에 이미 악한 준칙이 채택되어 있음을 보여 주는 것이다. 유혹에 굴복하는 의지의 허약성은 이미 도덕적 악을 전제로 한다.

그렇다면 우리는 악으로의 세 단계의 성벽 또한 마찬가지로 이미 도덕

51) ibid. p. 178.

적으로 타락한 인간의 사고 과정에 대한 기술로 볼 수 있지 않을까? 이러한 해석을 지지하는 문헌적 증거는 찾을 수 없으나 이러한 해석은 ‘악한 마음씨’와 ‘악으로의 성벽’ 사이의 충돌을 해소할 수 있다는 장점을 가지고 있다. 이 해석을 바탕으로 둘의 관계는 다음과 같이 다시 이해될 수 있다.⁵²⁾

심정의 허약성이라는 성벽을 가진 사람의 준칙을 재구성해 보자. 그는 자기사랑의 동기가 아주 강할 경우 도덕법칙을 쫓지 않아도 된다는 준칙을 가지고 있다. 이 준칙은 자기사랑의 동기를 도덕법칙의 동기보다 우위에 두는 궁극적 준칙, 즉, 악한 마음씨의 결과라 볼 수 있다. 그의 악이 심정의 허약성이라고 불리는 것은 그와 같은 선택을 한 주체가 자신의 행위의 책임을 온전히 자신에게 지우지 않기 위해 그 원인을 나약함으로 돌리고 그 결과 자신의 행위가 악한 행위임을 인지하는 데 실패한 사고 과정을 가리키는 표현일 뿐이다. 자신의 악함의 근원을 온전히 인식하는데 실패한 주체는, 그 결과 여전히 자신이 허약할지언정 궁극적으로 선하다는 기만에 빠지게 된다.

허약한 인간은 살아가면서 그를 도덕법칙에서 일탈케 하는 수많은 유혹들을 맞보게 되고 그럴수록 그의 안에서 자기사랑은 동기로서의 힘을 키워가게 된다. 이윽고 그는 도덕법칙을 최상의 동기로 두면서도 자기사랑의 동기가 충족될 때에만 그에 따르는 심정의 불순성의 단계에 도달하게 된다. 그런데 도덕법칙 준수의 조건이 자기사랑의 동기라면 실제로 그가 최상의 동기로 두고 있는 것은 도덕법칙이 아니라 자기사랑의 동기이다. 그렇다면 허약성과 마찬가지로 불순한 심정을 가진 사람이 도덕법칙을 최상의 동기로 둔다는 것은 사실에 있어 그러한 게 아니라 그 자신의 자기기만에 있어 그러한 것이다. 그는 실제로는 자기사랑의 동기를 최고의 동기로 두고 있으면서도 마치 그렇지 않은 척 자신을 기만하며, 그런 한에서 완전히 선하지는 않을지 몰라도 어느 정도는 선하다고 생각한다.

자기사랑은 이와 같이 악한 주체 안에서 자신의 정당성을 더욱 강하게 주장하기 시작한다. 이윽고 그는 명시적으로 자기사랑의 동기를 도덕법

52) 악이 단계적으로 심화되는 이 기본 구도는 Allison(1990) pp. 159-161의 해석을 참고하였다.

칙의 우위에 둔다. 그러나 그러면서도 그는 자신이 악하다고 생각하지 않는다. 오히려 자신의 행위가 합법칙적인 한에서 실제로는 자기 자신을 위해 한 행위임에도 불구하고, 자신을 선하다고 생각한다. 칸트가 말하고 있는 바처럼 만약 그가 운이 좋아 그의 상황이 자기사랑의 동기를 도덕법칙의 명령을 준수하면서 충족시키는 것을 허용한다면 그는 자신의 행위가 전적으로 합법칙적인 한에서 자기 자신을 완전히 선한 존재라 생각하는 것이다.

이러한 재구성은 심정의 허약성과 불순성이 도덕적인 악임을 분명히 해줌과 동시에 어떻게 악으로의 성벽이 허약성에서 전도성의 단계로 발전해 가는지를 설명해 준다. 그리고 이 재구성에 따르면 ‘악한 마음씨’와 ‘악으로의 성벽’은 단지 논의 대상의 범위에서만 다른 것이 아니다. 이 두 개념은 완전히 다른 도덕적 사태를 논하고 있다. ‘악한 마음씨’가 인간의 근본적인 도덕적 성질을 논하는 것이라면 ‘악으로의 성벽’은 악한 인간이 자기 자신의 도덕성에 대해 표상하는 바를 논하는 것이다.

또한 이러한 재구성에 근거해 ‘비사교적 사교성’의 발현에 대해 우드와는 다른 설명을 구성해 볼 수 있다. 우드는 우리가 타인의 시선에서 가치를 찾으려하기 때문에 비사교적 사교성을 발현시킨다고 주장한다. 그러나 인간의 가치는 오직 도덕적 가치일 수밖에 없으므로 타인의 시선에서 가치를 찾는 것은 자신의 도덕적 가치를 다른 환상적인 가치로 대체하는 것이다. 우드는 이와 같은 환상적 대체가 타인과의 접촉을 통해 마치 자연 발생적으로 일어나는 것처럼 서술하고 있다. 그러나 앞서 살펴본 것처럼 이는 악의 귀책성의 문제를 불러일으킨다. 따라서 우리는 우드와는 반대로 도덕적 가치를 다른 가치로 대체하려는 이러한 성향이 이미 악을 전제로 하는 것이라고 생각해야 한다. 인간이 자신의 도덕적 가치를 다른 것으로 대체해야만 하는 이유는 그가 어떠한 도덕적 가치도 갖고 있지 않기 때문이다. 도덕적으로 악한 주체가 자신을 기만하여 자신을 여전히 선한 사람이라 생각하게 되는 것처럼, 도덕적으로 악한 주체는 자신의 가치를 도덕성이 아니라 타인의 우위에 서는 것에서 찾으려고 한다. 이러한 생각의 계열을 따라 우리는 질투, 소유욕, 지배욕으로

상승되어 가는 비사교적 사교성의 단계들을 자기기만에서 출발하여 재구성해 볼 수 있을 것이다. 그러나 이 때 비사교적 사교성은 악의 궁극적 근원이라 할 수 없을 것이다. 그것은 특정한 “문명의 폐악”들의 근원일 수는 있으나 악 일반의 근원은 아닐 것이다.

이처럼 악한 마음씨를 자기기만의 성격을 갖는 것으로 이해함으로써 우리는 악을 향한 성벽 및 그 다양한 단계, 그리고 비교적 사교성으로서의 악까지 모두 그 자신 하나의 악한 행위의 결과라 할 수 있는 악한 마음씨로 소급하여 설명할 수 있다.

남은 과제는 개인의 악한 마음씨에 대한 논의가 어떻게 인류의 악으로의 성벽에 대한 논의로 연결되는지를 보이는 것이다. 인간 중에 악한 마음씨를 가진 사람이 있다는 것은 악한 행위가 있다는 것만큼이나 분명한 사실이다. 따라서 악한 마음씨를 가진 인간의 존재로부터 인류의 악으로의 성벽을 끌어 낼 수 있다면, 악의 보편성을 증명할 수 있다.

3.2. 악의 보편성에 대한 요청

칸트에 따르면 우리는 “자기 자신의 준칙마저도 언제나 관찰할 수 있지 않다.”(6:20) 따라서 우리 자신의 도덕적 가치는 우리에게 불투명한 것으로 남아 있다. 만약 우리가 언제나 우리 자신의 준칙을 명료하게 알 수 있다면 우리는 우리 자신이 악한 것을 바로 알 것이고, 그 누구도 “도덕법칙을 이룰테면 반역적으로(즉 불복종을 선언하면서) 포기하지는 않기 때문에”(6:36) 자기 자신이 도덕법칙을 거스르고 있음을 아는 순간 인간은 그렇게 행위하기를 그만둘 것이다. 인간에게 자기 자신의 준칙이 투명하게 드러나지 않기 때문에 인간이 자기 자신의 도덕적 가치를 기만할 위험이 생긴다. 실제로는 도덕법칙에 대한 순전한 존경에서 행위하고 있는 것이 아님에도 불구하고 자신이 그렇게 하고 있다고 생각할 수 있는 것이다. 만약 우리가 굉장히 호의적인 상황에 처해 있어 도덕법칙을 따르는 것이 우리의 자기사랑의 동기들을 저버릴 것을 요구하지 않을

때, 우리는 단 한 번도 의지의 나약함을 경험해 보지 않을지도 모르며, 그 결과 자기 자신을 도덕적으로 선하고 의지가 굳건한 사람으로 생각할 수도 있다. 그러나 이러한 상태에 처해 있는 사람은 상황이 변하면 언제든지 악한 행위를 할 수 있을 것이므로 실제로 그는 예지적으로는 악할지도 모르는 것이다. 이처럼 인간은 자신의 상태에 대해서 결코 확신할 수 없다. 자기 자신이 합법칙적인 행위를 하는 것이 도덕법칙에 대한 존경 때문인지 아니면 상황이 허락하기 때문인지 결코 확정을 내릴 수 없는 것이다. 게다가 경험은 오히려 우리 자신이 악할 가능성이 훨씬 높다는 사실을 말해 주고 있다. 역사와 인간 사회가 보여 주는 수많은 예들은 세상에는 악한 인간이 굉장히 많다는 것을 보여 주고 있으며, 합법칙적인 행위만을 하는 사람이라고 하여 결코 악하지 않다고 말할 수 없다는 것까지 고려하면 우리 자신이 선할 사람일 확률은 더욱 낮아진다.

그렇다면 상황 여하에 따라 등장할지도 모르는 악을 방지하기 위해서 그가 할 수 있는 일은 자기 자신을 이미 악하다고 생각하고, 따라서 자기 자신을 선을 향해 부단히 노력해야만 하는 존재로 생각하는 것 외에 다른 것일 수 없다. 그렇지 않고 그가 자기 자신을 선하다고 생각할 경우, 그가 경향성과 도덕법칙의 요구가 충돌하는 상황에 놓여 악한 행위를 하게 되었을 때 그는 여전히 자기기만을 통해 변명거리를 만들어 내게 될 것이다. 그렇다면 이미 그는 도덕적인 개선의 가능성에서 상당히 떨어져 있는 것이다. 따라서 발생 가능한 악을 방지하고, 도덕적 개선 가능성을 높이기 위해서는 자신을 악하다고 생각해야 하는 것이 더 낫다. 그리고 악을 방지하고, 도덕적 개선 가능성을 높이기 위해 할 수 있는 일을 다 하는 것은 우리의 도덕적 의무이므로 우리는 자신을 악한 사람이라고 생각해야 한다. 그리고 그 누구도 자신의 선하다는 것을 확신할 수 없기 때문에, 우리 모두는 자신을 악한 사람이라고 생각해야 한다. 우리 모두가 우리 자신을 악한 사람이라고 생각해야만 하는 이상, 우리 모두는 악하다.

물론 이것은 악의 보편성에 대한 칸트의 주장을 충분히 뒷받침해 주지 않는다. 물론 우리가 우리 자신을 악하다고 생각해야 하는 이상, 그 누

구도 자기 자신을 선한 인간의 예로서 들어 보일 수 없으므로, 이 주장은 그 어떤 경험적인 반박에 대해서도 안전하다고 할 수 있다. 선한 사람을 단 하나만 반례로 들어보라고 요구하며, 인간 악성의 보편성을 주장했을 때 이에 반박할 수 있는 사람은 없다. 칸트에게는 자기 자신을 악하다고 생각할 충분한 이유가 있었다. 또한 칸트의 글의 모든 잠재적 독자-즉, 인류-도 자기 자신을 악하다고 생각할 충분한 이유가 있다. 그러나 모든 사람이 자기 자신을 악하다고 생각한다는 것과 모든 사람이 악하다는 주장은 다르다. 어떠한 반증도 존재할 수 없다는 점에서 이 주장은 강력하지만 이 주장은 그 자체의 근거를 가지고 있지 않으며, 따라서 칸트의 주장은 철학적 논변이라기보다 일종의 수사학적 전략으로 이해될 수 있다. 즉, 칸트의 주장은 악의 보편성에 대한 어떤 실제적 증명을 위한 것이 아니라 인류를 선을 향해 나아가게 하기 위한 일종의 자극제로서 제기된 것으로 보아야 한다는 것이다. 즉, 이 증명은 보편적으로 타당한 진리가 아니다. 악하지 않을 사람이 실제로 존재할 수도 있다. 그러나 우리는 경험을 통해서도 결코 이를 알 수 없다. 따라서 최소한 칸트의 독자에게는 악의 보편성 주장은 반증할 수 없는 것이며, 우리의 도덕적 삶을 위해서 받아 들여야만 하는 실천적 진리이다.

5. 결론

유한한 존재인 인간은 이성의 명령을 의식하는 동시에 감성적 욕구 역시 가지고 있으며, 인간에게 이성의 요구와 감성의 요구는 동일한 힘을 가지고 있는 것으로 여겨진다. 이러한 인간의 근본적인 실존적 조건은 인간을 언제나 악의 상태로 전락할 위험에 처해 있게 한다. 그러나 인간이 자유로운 존재인 한, 인간의 선과 악은 이러한 조건에 의해 필연적으로 결정되는 것이 아니라 자유로운 선택에 의해 우연적으로 결정되는 것이다. 인간은 도덕적으로 선하거나 악한 것 중 어느 하나 일 수밖에 없

기 때문에 이 선택은 경험세계에서 인간의 모든 자유로운 행위 이전에 행해진 것이라 할 수 있다. 그러므로 경험 세계에서 인간의 도덕적 성질은 이미 결정되어 있지만, 인간은 예지적인 행위를 통해 자신의 도덕적 성질을 창시한다는 의미에서 궁극적으로 자유로운 존재이다. 그럼에도 불구하고 모든 인간은 자유롭게 악하게 될 것을 선택하였기에 근본적으로 악하다.

이 논문은 이와 같이 쉽게 이해하기 어려운 칸트의 주장을 해명하는 것을 목표로 하였다. 이 논문을 통해 해명된 것은 다음과 같다.

인간의 악성은 자기기만을 기본적인 성격으로 갖는다. 많은 경우 우리는 자신이 범한 잘못의 원인을 유혹이나 의지의 나약함으로 돌림으로써 자신의 책임을 회피하고 도덕적 개선을 위한 진지한 노력을 기울이지 않는다. 칸트에 따르면 우리가 그와 같은 행동을 하는 이유는 바로 우리 자신이 이미 악하기 때문이다. 우리가 우리 자신을 속여 자신의 잘못을 직시하지 못하게 하는 것이다. 『종교』에서 소개된 ‘악한 마음씨’와 ‘악으로의 성벽’ 개념 사이에는 중복되는 내용과 모순되는 내용이 존재하는데 우리는 이 문제를 ‘악한 마음씨’를 우리 자신의 근본적인 악함으로 이해하고, ‘악으로의 성벽’을 ‘악한 마음씨’를 가진 우리가 자기기만의 결과 갖게 되는 자기 이해로 이해함으로써 해결할 수 있다.

‘악으로의 성벽’을 기만적인 자기 이해로 봄으로써 우리는 왜 모든 인간이 이러한 성벽을 갖고 있다고 전제해야 하는지를 이해할 수 있다. ‘악으로의 성벽’의 핵심은 자기 자신을 선하다고 생각하는 기만에 있다. 우리의 경험이 보여 주는바 세상에 악이 그렇게 만연해 있다면 우리가 자신을 선하다고 생각할 때 사실 우리는 이미 악하며 자신을 기만하고 있는 것일 수 있다. 이러한 자기기만의 위험성은 도덕적 개선의 여지를 없애 버린다는 것이다. 자기기만에 빠져 스스로를 선하다고 생각하고 있는 한 인간은 도덕적 개선을 위한 노력을 기울일 수 없기 때문이다. 따라서 인간은 스스로를 악하다고 생각하고, 여기서 출발하여 선을 실현하도록 노력해야 한다.

이 논문은 이상과 같이 ‘악한 마음씨’와 ‘악으로의 성벽’ 개념을 명료

화함으로써 모든 인간이 근본적으로 악하다는 칸트의 증명되지 않은 주장에 대한 대안적 증명을 구성하였다. 그러나 이 논문은 칸트 실천철학의 대중을 이루는 선의지 개념과 근본악의 관계에 대해서 해명하지 못하였다는 한계를 가지고 있다. 칸트의 근본악 테제가 그토록 큰 스캔들을 불러일으킨 것은 이 테제가 인간의 선의지에 대한 칸트의 이전 주장과 충돌하는 것처럼 보이기 때문이었음을 생각하면 이는 중대한 문제이다. 선의지와 근본악의 관계에 대해 체계적인 해명을 하는 것은 이후의 연구 과제로 남겨 두고, 여기서는 두 개념이 서로를 배제하는 것은 아니라는 점을 간단히 밝히는 것으로 논의를 마무리하겠다.

칸트는 『종교』에서 인간이 근본적으로 악하다는 주장을 하면서도 인간이 선한 존재가 될 가능성을 부정하지 않았다. 이는 인간이 근본적으로 악함에도 불구하고, 인간 안의 선의지가 근절되는 일은 없다는 것을 의미한다. 인간이 근본적으로 악하지만 동시에 선의지 또한 갖고 있다는 것이 어떻게 가능한가? 이 논문에서 인간의 악성은 인간이 실제로 악한지의 여부에 있어서 증명된 것이 아니라 모든 인간이 도덕적으로 선행하기 위해서는 자신을 어떻게 생각해야 하는가에 있어서 증명되었다. 자기 자신이 악하다는 인간의 도덕적 자기 이해는 인간의 선의지에 의해 요구되는 것이다. 이처럼 근본악이 선의지를 전제로 도출되고 있다는 점에서 근본악 테제와 선의지는 서로를 배제하지 않는다. 선의지와 근본악 모두 인간의 실천적 삶에서 요청되는 개념이며, 인간의 선의지는 무조건적으로 요청되는 반면, 근본악은 선을 추구하기 위해 조건적으로 요청되는 것이란 점에서 위상이 다른 두 개념이다. 그러나 인간의 도덕적 성격을 판정할 때 칸트는 선의지 개념에 의거하지 않고, 근본악 개념에 의거한다. 근본악과 선의지의 관계를 해명하기 위해서는 왜 인간의 도덕적 성격이 전자에 의해 판정되어야만 하는지를 밝혀야 할 것이다.

참고 문헌

1. 1차 문헌

Kritik der reinen Vernunft[KrV], AA III-IV.

『순수이성비판』, 백종현 옮김, 아카넷, 2006.

Kritik der praktischen Vernunft[KpV], AA V

『실천이성비판』, 백종현 옮김, 아카넷, 개정판 2009.

Grundlegung zur Metaphysik der Sitten, AA IV

『윤리형이상학 정초』, 백종현 옮김, 아카넷, 2005.

Idee zu einer allgemeinen Geschichte in weltbürgerliche Absicht, AA VII.

『칸트의 역사철학』, 이한구 옮김, 서광사, 개정판 2009.

Die Metaphysik der Sitten[MS], AA VI.

『윤리형이상학』, 백종현 옮김, 아카넷, 2012.

Die Religion innerhalb der Grenzen der bloßen Vernunft, AA VI.

『이성의 한계 안에서의 종교』, 백종현 옮김, 아카넷, 2011.

Eine Vorlesung Kant's über Ethik im Auftrage der Kantgesellschaft.

Lectures on Ethics, trans. by. L. Infield, Hackett Publishing Company, Indianapolis-Cambridge, 1980.

2. 2차 문헌

Allison, H. E., *Kant's theory of Freedom*, Cambridge University Press, 1990.

Wood, Allen W., *Kantian Ethical Thought*, Cambridge University Press, 1999

-----, "Kant and the Intelligibility of Evil", *Kant's Anatomy of Evil*, Cambridge University Press, 2009.

-----, *Kant's Moral Religion*, Cornell University Press, 1970.

Muchjik, Pablo, "An Alternative Proof of the Universal Propensity to Evil", *Kant's Anatomy of Evil*, Cambridge University Press, 2009.

Jeanine M. Grenberg, "Social Dimensions of Kant's Conception of Radical Evil", *Kant's Anatomy of Evil*, Cambridge University Press, 2009.

Rossi, P. J. , *The Social Authority of Reason, Kant's Critique, Radical Evil, and the Destiny of Humankind*, State University of New York Press, 2005.

Sussman, David, "Perversity of the Heart", *Philosophical Review*, Vol 114, No.2, Ducke University Press. 2005.

Grimm, Stephen R., "Kant's Argument for Radical Evil", *European Journal of Philosophy* 10:2, Blackwell, 2002.

Sidgwick, Henry, "The Kantian Conception of Free will," *Mind* 1, 1888.

Duncan, Samuel, *Kant on Freedom, Reason, and Moral Evil*, University of Virginia, 2011.

O'Neill, Onora, *Kant on Reason and Religion*, University of Utah Press, 1997.

신옥희, 「칸트에 있어서 根本惡과 神 : 「宗教論」에 나타난 道德的 神存在 證明」, 『哲學』, Vol. 18, 1982.

김종식, 「칸트와 악의 문제」, 『大同哲學』, Vol.28, 2004.

Abstract

A Study on Kant's concept of Radical Evil

LEE, Duckkyun

Department of Philosophy
the Graduate School
Seoul National University

In his “Religion within the Boundaries of Mere Reason” Kant introduced a very scandalous concept, Radical Evil, which argues that man, having chosen the inextirpable “propensity to evil” (Hang zum Bösen), is evil by nature. This thesis impugned by many of his contemporaries due to its seemingly compromising attitude toward a christian dogma which condemns humanity for being born with original sin. Therefore, the thesis was not considered to be meaningful in terms of Kant's philosophy. In addition, the absence of the verification of the Thesis made it difficult for scholars to consider it as deserving careful reading.

However, as studies over Kant's late works have been carried out for last few decades, a variety of efforts to read the thesis in the context of Kant's whole philosophical project are flourishing. These efforts focus on filling the lack of verification of the Thesis and can be divided into two different readings according to their strategies of proving the Radical Evil thesis. One type, transcendental reading, understands Kant's thesis as synthetic a priori judgment and tries to give a transcendental deduction for it. The other type, anthropological reading, comprehends Kant's thesis as having anthropological

implication and tries to find an element in human being that constitutes human nature and makes man evil.

Although these two readings gave many significant insights in understanding the Thesis, the two readings fails to give an account about how evil can be imputable to man if he is evil by nature. Since man can not be judged as morally good or evil if he is not responsible for his moral state in Kant's philosophy, this failure is crucial. Aside from solving this problem of responsibility for evil, no alternative argument for the thesis can be regarded successful.

The present study aims to give another argument which can solve the problem of responsibility for evil. For this purpose, I focus on the concept of "evil disposition" (*Böse Gesinnung*) and the concept of "propensity to evil" (*Hang zum Bösen*). These two concepts seem to have a similarity because both of them define what evil is in a same way explaining that man is evil in that she adopts a maxim in which the motive of duty is subordinate to the motive of self-love. The former offers this explanation in terms of individual's evil and the latter humanity's evil. Therefore, it seems that two concepts are different only as to its scope of discussion. However, there is a crucial difference among the two concepts. The former denies that the origin of evil can be traced further. The origin of man's evil can not be other than our choice. The latter, however, traces the origin of evil back to "the fragility of the heart." (*Gebrechlichkeit des Herzen*) It suggests that we should consider these two concepts different indicating different moral states, respectively.

Based on the Kant's explanation that evil makes us deceive ourselves about our moral character, I concluded that "propensity of evil" indicates our self-deceptive moral state that we came to have because of our "evil disposition." To put it simply, a morally evil person whose disposition is evil by his own choice develops "propensity to evil" which makes her think that she is morally good

despite of her evil disposition.

Given this understanding, we can induce an alternative argument for the Radical Evil thesis. If evil has this deceptive character, it could be always the case that we are evil at the very moment when we consider ourselves good. Hence, it is better for us to consider ourselves morally evil to prevent us from further moral degeneration and to promote good. Also, no one can assert that he is good, there can be no counterexample of the Radical Evil Thesis. In addition, the prevalence of evil all around the world and human history hitherto is so obvious; it is plausible to contend that man is evil by his own fault. Since this argument can not be refuted and we should be careful about our self-deceptive moral nature, it is reasonable to accept the Thesis. Consequently, the Radical Evil thesis can be proved practically, not theoretically.

Key words: Radical Evil, morality, moral value, responsibility for evil, evil disposition, propensity to evil, self-deception

Students Number: 2009-22773